

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2001, 09:38:08 ; Search time 464.47 Seconds  
(without alignments)  
11586.851 Million cell updates/sec

Title: US-09-522-753-4  
Perfect score: 8571  
Sequence: 1 catgtcgggctccacacagc.....caaaaaaaaaaaaaaaaaaaaaa 8571

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	8495.2	99.1	8564	21	Human ORFX ORF338
2	949.4	11.1	1027	21	Human pancreatic c
3	898.4	10.5	956	21	Human proliferatio
4	898.4	10.5	956	21	Human proliferatio
5	629.6	7.3	7780	21	HNCR nucleotide s
6	626.8	7.3	7900	21	Human HNRKR encodi
7	409.4	4.8	427	22	Novel human polynu
8	376	4.4	872	21	Human colon cancer
9	311.2	3.6	312	21	Human secreted pro
10	206.2	2.4	226	16	Human gene signatu
11	187.8	2.2	789	20	Human gene express

	12	155	1.8	300	21	AAA01573	Human colon cancer
	13	142	1.7	1922	17	AAF31931	Retinoid X recepto
	14	130.8	1.5	342	22	AAF67531	Novel human polynu
	15	118.4	1.4	114955	20	AA53491	Human adenosine A1
C	16	107.8	1.3	3489	21	AA530290	Kaposi's sarcoma-a
	17	107.8	1.3	32207	20	AAV73805	KSHV LUR DNA (nucl
C	18	107.8	1.3	137507	19	AAV19941	KSHV long unique c
	19	98.8	1.2	1337	20	AA217263	Human gene express
	20	97	1.1	114955	20	AA53491	Human adenosine A1
	21	96.2	1.1	799	19	AAV55831	Nucleotide sequenc
	22	96.2	1.1	1926	21	AA50254	Epstein Barr virus
	23	96.2	1.1	2580	21	AA75454	Nucleotide sequenc
C	24	96.2	1.1	5452	20	AA90923	Anti-sense strand
C	25	96.2	1.1	8705	20	AA23778	Vector pshuttle DN
	26	96.2	1.1	9600	19	AAV21683	Vector plasmid pCM
	27	96.2	1.1	10380	20	AA22248	Nucleotide sequenc
	28	96.2	1.1	10596	14	AAQ51731	Plasmid pCisEBON f
	29	96.2	1.1	10596	17	AA740348	Plasmid pCisEBON f
	30	96.2	1.1	10596	20	AA15650	Nucleotide sequenc
C	31	96.2	1.1	16080	21	AA55953	DNA clone pCEK Cl.
	32	93	1.1	94	21	AA25194	Human secreted pro
	33	91.8	1.1	795	19	AAV55830	FLGA insert stabl
	34	91.6	1.1	2004	18	AA785356	Nephila clavipes s
C	35	88.4	1.0	1000	21	AA02484	Human colon cancer
	36	86.2	1.0	1995	12	AAQ14184	N.clavipes draglin
	37	86.2	1.0	1995	19	AAV23250	Nephila clavipes s
	38	86.2	1.0	1995	21	AA238196	N. clavipes spider
	39	85.6	1.0	1925	20	AA90924	Epstein Barr Virus
	40	85.4	1.0	16442	18	AA83006	Partial mouse WRN
C	41	83.4	1.0	144	16	AA722542	Human gene signatu
	42	81.6	1.0	2338	12	AAQ14183	N.clavipes draglin
	43	81.6	1.0	2338	19	AAV23249	Nephila clavipes s
	44	81.6	1.0	2338	21	AA238195	N. clavipes spider
	45	81.4	0.9	9551	20	AAZ22301	CDNA encoding a hu

ALIGNMENTS

RESULT 1  
AAC74783  
ID AAC74783 standard; cdna; 8564 BP.  
XX AC AAC74783;  
XX AC AAC74783;  
DT 08-FEB-2001 (first entry)  
XX DE Human ORFX ORF338 polynucleotide sequence SEQ ID NO:675.  
XX DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; aschma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.















Qy 5879 ctgcgaagccccagccgctcgggctgagccgctcctcccccagcaaggctcg 5938  
Db 5881 ctgcgaagccccagccgctcgggctgagccgctcctccccagcaaggctcg 5940  
Qy 5939 gagcccgccctagtgctctctctctgagccacgcccacatcgcccgcacccctcg 5998  
Db 5941 gagcccgccctagtgctctctctctgagccacgcccacatcgcccgcacccctcg 6000  
Qy 5999 aagaacctcgacctcaccagccagccgagcccgccgcccgcacacctgctcgccctcg 6058  
Db 6001 aagaacctcgacctcaccagccagccgagcccgccgcccgcacacctgctcgccctcg 6060  
Qy 6059 gaccgcacccgggaaagactcaagtaaaccttttccatccaggaaactggaactccgt 6118  
Db 6061 gaccgcacccgggaaagactcaagtaaaccttttccatccaggaaactggaactccgt 6120  
Qy 6119 tctctgggttaccagcgagctacagcccgcaaggggtgagcccgctgagcccttg 6178  
Db 6121 tctctgggttaccagcgagctacagcccgcaaggggtgagcccgctgagcccttg 6180  
Qy 6179 agctcaccagctgacccacgacaaggggtccccaagcaccctggaagagctcgacaag 6238  
Db 6181 agctcaccagctgacccacgacaaggggtccccaagcaccctggaagagctcgacaag 6240  
Qy 6239 agccacctggaggagctcgcccaagcagcagcccgctgaaactctggcgggag 6298  
Db 6241 agccacctggaggagctcgcccaagcagcagcccgctgaaactctggcgggag 6300  
Qy 6299 gccgccaacctccacacctggccgctgctgagagccagccctcgtaagcccgctg 6358  
Db 6301 gccgccaacctccacacctggccgctgctgagagccagccctcgtaagcccgctg 6360  
Qy 6359 ctccagacccgcccaggggtcaaaagttcaccagcggtggttcacctggcccagcacatc 6418  
Db 6361 ctccagacccgcccaggggtcaaaagttcaccagcggtggttcacctggcccagcacatc 6420  
Qy 6419 agtgaggctcaacacaggaatacaccggaacacccacacagcagctcagcgacccctg 6478  
Db 6421 agtgaggctcaacacaggaatacaccggaacacccacacagcagctcagcgacccctg 6480  
Qy 6479 cccgccccctctactctccctggggccagctgcccgtctctggacctccgcgccca 6538  
Db 6481 cccgccccctctactctccctggggccagctgcccgtctctggacctccgcgccca 6540  
Qy 6539 cccagtgacctctactctccgcccggaccatggtgcccgcccgctggctccccccac 6598  
Db 6541 cccagtgacctctactctccgcccggaccatggtgcccgcccgctggctccccccac 6600  
Qy 6599 agcgaaaggggcaagaggtctccagagccaaacagacgctcgttggtggtgag 6658  
Db 6601 agcgaaaggggcaagaggtctccagagccaaacagacgctcgttggtggtgag 6660  
Qy 6659 gacgggtatgaaactgtgtcccccagggagcatgacggagcaggcactcccgaggt 6718  
Db 6661 gacgggtatgaaactgtgtcccccagggagcatgacggagcaggcactcccgaggt 6720  
Qy 6719 gctgtgtaccgctgctgtaccgggatgggaacagacgagcccgagatggctcc 6778  
Db 6721 gctgtgtaccgctgctgtaccgggatgggaacagacgagcccgagatggctcc 6780  
Qy 6779 aagttccaggaacacacagccagccgacgcttcttcagaaactgacgagagcaac 6838  
Db 6781 aagttccaggaacacacagccagccgacgcttcttcagaaactgacgagagcaac 6840  
Qy 6839 tccgcatggttcaagtccaagaagaagatcaacaagaagctgaacacccacaacccg 6898  
Db 6841 tccgcatggttcaagtccaagaagaagatcaacaagaagctgaacacccacaacccg 6900  
Qy 6899 aatgagctgaatacaatatcagccagcctggagcgagatcttcaatatgcccgcac 6958  
Db 6901 aatgagctgaatacaatatcagccagcctggagcgagatcttcaatatgcccgcac 6960  
Qy 6959 accggaacaggccttatgacctatagaagccaggcggtgaggaacatgcccagcacaac 7018

Db 6961 accggaacaggccttatgacctatagaagccaggcggtgaggaacatgcccagcacaac 7020  
Qy 7019 atggggctgagggccataattagaaaagcactcatggttaaatatgaccagtggaagag 7078  
Db 7021 atggggctgagggccataattagaaaagcactcatggttaaatatgaccagtggaagag 7080  
Qy 7079 tcccgcgctcagcgccaatgcttttaacctctgaatgccagtgccagctgcccct 7138  
Db 7081 tcccgcgctcagcgccaatgcttttaacctctgaatgccagtgccagctgcccct 7140  
Qy 7139 gctatgcccataaccgctgctgacggagcgagtgaccacacactcactcgccaggtgac 7198  
Db 7141 gctatgcccataaacgctgctgacggagcgagtgaccacacactcactcgccaggtgac 7200  
Qy 7199 ggccggaagcgcaaggtctctgagcagaccagcagccgaaagcacaagtcgcccgcccg 7258  
Db 7201 ggccggaagcgcaaggtctctgagcagaccagcagccgaaagcacaagtcgcccgcccg 7260  
Qy 7259 ggctggcatctgggagccgcccacctctgtctcctcagtgactcgagggagactgc 7318  
Db 7261 ggctggcatctgggagccgcccacctctgtctcctcagtgactcgagggagactgc 7320  
Qy 7319 aaccgcccagcgccctcaccacccgctgtgtggagagcaggccctcgccaggttcc 7378  
Db 7321 aaccgcccagcgccctcaccacccgctgtgtggagagcaggccctcgccaggttcc 7380  
Qy 7379 acgcatctccctcaacaacccctgatcatcgctgagcggtgtgatgcttcccca 7438  
Db 7381 acgcatctccctcaacaacccctgatcatcgctgagcggtgtgatgcttcccca 7440  
Qy 7439 ccccacccgggctcccgccgagcgccctcgctggccccacacacccctgggac 7498  
Db 7441 ccccacccgggctcccgccgagcgccctcgctggccccacacacccctgggac 7500  
Qy 7499 gaggagcccaagcactgctgtctgcagtaagagacactctcgacagcgagtgactc 7558  
Db 7501 gaggagcccaagcactgctgtctgcagtaagagacactctcgacagcgagtgactc 7560  
Qy 7559 aqaacagcgcgggggggcggtggttctgaggtcccgagcgagccacag 7618  
Db 7561 aqaacagcgcgggggggcggtggttctgaggtcccgagcgagccacag 7610  
Qy 7619 aacggccctgagggagcgggcgctgcgactcccccaacaaaggagccctgag 7678  
Db 7611 aacggccctgagggagcgggcgctgcgactcccccaacaaaggagccctgag 7670  
Qy 7679 tccgctcgctccatccatctgtccgtccagagccggcatccttgcgtctaaagcc 7738  
Db 7671 tccgctcgctccatccatctgtccgtccagagccggcatccttgcgtctaaagcc 7730  
Qy 7739 ttaactaaagactcccccgccgctggcctgtgcagacttactcaggggatgttacc 7798  
Db 7731 ttaactaaagactcccccgccgctggcctgtgcagacttactcaggggatgttacc 7790  
Qy 7799 tggctcggggaagggagggcgggcgagggggcgagggcgagcggtgtggcagcc 7858  
Db 7791 tggctcggggaagggagggcgggcgagggcgagggcgagcggtgtggcagcc 7850  
Qy 7859 acacacagggcccgggcgccagggagcccaagcaggatgacacgacacctccagcc 7918  
Db 7851 acacacagggcccgggcgccagggagcccaagcaggatgacacgacacctccagcc 7910  
Qy 7919 actgctcccccgatgatttgaaacaaagtctaaactgagctcgagcccccgcc 7978  
Db 7911 actgctcccccgatgatttgaaacaaagtctaaactgagctcgagcccccgcc 7970  
Qy 7979 ctccctcgctcccatcccgcttagcgctctggacagatggacgagccctgtccagc 8038  
Db 7971 ctccctcgctcccatcccgcttagcgctctggacagatggacgagccctgtccagc 8030  
Qy 8039 cccagtgctgcttcgggtcccccagactgccccagcactgagattgtgtgaaacc 8098







QY 8183 catctgtccgcagcggtgggaagaaagcagatgtaaatgatgtgtgtttacagg 8242  
 Db |||||||  
 QY 603 catcgttgcgcagcggtgggaagaaagcagatgtaaatratgtgtgtttacagg 662  
 Db |||||||  
 QY 8243 tataattttgataccctcaatgaataattcagatgttttaacgaaggaactaccc 8302  
 Db |||||||  
 QY 663 tataattttgataccctcaatgaataattcagatgttttaacgaaggaactaccc 722  
 Db |||||||  
 QY 8303 agtattactgctgtgtcttttctctctgtcttaccgttcaagagcggtgtgcaggcg 8362  
 Db |||||||  
 QY 723 agtattactgctgtgtcttttctctctgtcttaccgttcaagagcggtgtgcaggcg 782  
 Db |||||||  
 QY 8363 acagtcggtgaccccatcactgcaggaaccaaggggcggaactgctcg-tcagccccc 8421  
 Db |||||||  
 QY 783 acagtcggtgaccccatcactgcaggaaccaaggggcggaactgctggtccacgcccc 842  
 Db |||||||  
 QY 8422 gctgtgtctccctccctccctctctcttcttgggcagaatgaattcgtattcttgccc 8481  
 Db |||||||  
 QY 843 gctgtgtctccctccctccctctcttcttgggcagaatgaattcgtattcttgccc 902  
 Db |||||||  
 QY 8482 gccatttgcgcaggggtgtgtattctgtcatttacacacgtcttcaattataaaagcg 8541  
 Db |||||||  
 QY 903 gccatttgcgcaggggtgtgtattctgtcatttacacacgtcttcaattataaaagcg 962  
 Db |||||||  
 QY 8542 aattatactccaaaaaaataaaataaa 8571  
 Db |||||||  
 QY 963 aattatactccaaaaaaataaaataaa 992  
 Db |||||||

## RESULT 3

AAA98862  
 ID AAA98862 standard; cDNA; 956 BP.

XX AC AAA98862;

XX DT 16-FEB-2001 (first entry)

XX DE Human proliferation-associated endothelial EST cDNA SEQ ID NO 8.

XX EST; expressed sequence tag; human; cell differentiation; antidiabetic;  
 KW cell proliferation; endothelial tissue; angiogenic; antipsoriatic;  
 KW ophthalmological; nephrotropic; cytostatic; hepatotropic; antiarthritic;  
 KW antiarteriosclerotic; antirheumatic; gene therapy; angiogenic disease;  
 KW treatment; rheumatoid arthritis; haemangioma; angiofibroma; eye disease;  
 KW diabetic retinopathy; glaucoma; kidney disease; glomerulonephritis;  
 KW diabetic nephropathy; malign nephrosclerosis; transplant rejection;  
 KW thrombotic microangiopathic syndrome; glomerulopathy; fibrotic disease;  
 KW mesangial cell proliferative disease; ss.

XX OS Homo sapiens.

XX PN DE19911684-A1.

XX PD 14-SEP-2000.

XX PF 09-MAR-1999; 99DE-1011684.

XX PR 09-MAR-1999; 99DE-1011684.

XX PA (SCHD ) SCHERING AG.

XX PI Glienke J, Thierauch K, Hinzmann B, Pillarsky C;

XX DR WPI; 2000-588417/56.

XX PT Human nucleic acid sequences from endothelial cells useful for  
 PT treatment of angiogenic disease, including psoriasis, arthritis and  
 PT diabetic retinopathy, etc -

XX PS Claim 1a; Page 63; 84pp; German.

XX This invention describes novel human nucleic acid sequences (I) from  
 CC endothelial cells which have angiogenic, antipsoriatic, antidiabetic,  
 CC

CC ophthalmological, nephrotropic, cytostatic, hepatotropic,  
 CC antiarteriosclerotic, antiarthritic, antirheumatic activity and can be  
 CC used for gene therapy. The nucleic acid sequences may find use as tools  
 CC to identify agents against angiogenic diseases. The polypeptides may be  
 CC used in medicaments for gene therapy to treat angiogenic diseases. (I)  
 CC in sense or antisense form may be used. The genomic genes may also be  
 CC used with suitable promoters and/or enhancers. The sequences (nucleic  
 CC acid and proteins) may be used to treat psoriasis, arthritis, especially  
 CC rheumatoid arthritis, haemangioma, angiofibroma, eye diseases, especially  
 CC diabetic retinopathy, neovascular glaucoma, kidney disease, such as  
 CC glomerulonephritis, diabetic nephropathy, malign nephrosclerosis,  
 CC thrombotic microangiopathic syndrome, transplantation rejection and  
 CC glomerulopathy, fibrotic disease, such as liver cirrhosis, mesangial cell  
 CC proliferative disease, arteriosclerosis and injury to nerve tissue.  
 XX

SQ Sequence 956 BP; 192 A; 288 C; 287 G; 189 T; 0 other;

Query Match 10.5%; Score 898.4; DB 21; Length 956;  
 Best Local Similarity 99.1%; Pred. No. 2e-144;  
 Matches 924; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 7627 tgcagagcgggcggtgcgactcccccacaaaggaagagccctgagtcgcctg 7686  
 Db |||||||

QY 25 tgcagagcgggcggtgcgactgcccacaaaggaagagccctgagtcgcctg 84  
 Db |||||||

QY 7687 cgcctccatccatctgcgtccagagcgcgcattctcctgtctaaagccttaactaa 7746  
 Db |||||||

QY 85 cgcctccatccatctgcgtccagagcgcgcattctcctgtctaaagccttaactaa 144  
 Db |||||||

QY 7747 gactcccgccccggcggtgcgacttctactcaggaggtgtttacctggtgtctc 7806  
 Db |||||||

QY 145 gactcccgccccggcggtgcgacttctactcaggaggtgtttacctggtgtctc 204  
 Db |||||||

QY 7807 gggagggaggggaaagggcggggagggggcacggcagcgctgtgtggcgacacacag 7866  
 Db |||||||

QY 205 gggagggaggggaaagggcggggagggggcacggcagcgctgtgtggcgacacacag 264  
 Db |||||||

QY 7867 gcggcgagggcgcgaggaacccaaagcagatgacacgcacccctccacgcactgcctc 7926  
 Db |||||||

QY 265 gcggcgagggcgcgaggaacccaaagcagatgacacgcacccctccacgcactgcctc 324  
 Db |||||||

QY 7927 ccccgaaatcatttggaaacaaagtctaaactgagctgcagcccccgccctccctcc 7986  
 Db |||||||

QY 325 ccccgaaatcatttggaaacaaagtctaaactgagctgcagcccccgccctccctcc 384  
 Db |||||||

QY 7987 gctcccatcccgcttagcgtctgtggacagatggacgcaggccctgtccagcccccaagt 8046  
 Db |||||||

QY 385 gctcccatcccgcttagcgtctgtggacagatggacgcaggccctgtccagcccccaagt 444  
 Db |||||||

QY 8047 cgcgtgttcggtcccccacagactgcccacgaacagagatgctggaaacaaagtcagg 8106  
 Db |||||||

QY 445 cgcgtgttcggtcccccacagactgcccacgaacagagatgctggaaacaaagtcagg 504  
 Db |||||||

QY 8107 ccaggtggcggaacaaagggcgaggtgcgctgcgggggaaacggagatgctcgagagactg 8166  
 Db |||||||

QY 505 ccaggtggcggaacaaagggcgaggtgcgctgcgggggaaacggagatgctcgagagactg 564  
 Db |||||||

QY 8167 gactgttttttccacacatcgttgcgcagcggtgggaaaggaagcagatgtaaatgat 8226  
 Db |||||||

QY 565 gactgttttttccacacatcgttgcgcagcggtgggaaaggaagcagatgtaaatgat 624  
 Db |||||||

QY 8227 gtgttggtttacaggggtatatttttgataccttcaatgaattacatgattttacgc 8286  
 Db |||||||

QY 625 gtgttggtttacaggggtatatttttgataccttcaatgaattacatgattttacgc 584  
 Db |||||||

QY 8287 aaggaaggaacttaccagattactgctgtgctgtgtttgattctctgttaccgttcaag 8346  
 Db |||||||

QY 685 aaggaaggaacttaccagattactgctgtgctgtgtttgattctctgttaccgttcaag 744  
 Db |||||||

QY 8347 aggcgtgtgcagggccgcagagtcggtgaccccatcactgcaggaacaaagggcggggac 8406  
 Db |||||||

QY 745 aggcgtgtgcagggccgcagagtcggtgaccccatcactgcaggaacaaagggcggggac 804  
 Db |||||||



















AAA00010 to AAA02716 represent polynucleotides isolated from cDNA libraries constructed from human colon cancer cell lines. The present invention also describes a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived. The polynucleotides sequences can be used in a method for detecting differentially expressed genes correlated with a cancerous state of a mammalian cell. The polynucleotides can also be used as probes for detecting and mapping related genes. They can be used in diagnosis and prognosis of diseases and disorders (e.g. identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). This is, particularly for breast cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-negative breast cancer, lung cancer, and colon cancer.

Sequence 872 BP: 345 A; 133 C; 203 G; 185 T; 6 other;  
xx  
sq



QY 1620 acgacaagaagactcctcaagagaagacagacacacacctcaggagacgaacgacg 1679  
|||||  
Db 61 acgacaagaagactcctcaagagaagacagacacacacctcaggagacgaacgacg 120  
|||||  
QY 1680 agaagagctgtgctcccaaaagcccaaaactgccaacagccagggaggaacgcaag 1739  
|||||  
Db 121 agaagagctgtgctcccaaaagcccaaaactgccaacagccagggaggaacgcaag 180  
|||||  
QY 1740 gccgatcacccgctcaatgcttaagggccaacagcagcagggagggccatcaccccccagc 1799  
|||||  
Db 181 gccgatcacccgctcaatgcttaagggccaacagcagcagggagggccatcaccccccagc 240  
|||||  
QY 1800 agagcgcgagctgctcccaatggagctgaatgagagttctcgtggacagagaagaaaa 1859  
|||||  
Db 241 agagcgcgagctgctcccaatggagctgaatgagagttctcgtggacagagaagaaaa 300  
|||||  
QY 1860 tggaaacagcca 1871  
|||||  
Db 301 tggaaacagcca 312  
|||||  
RESULT 10  
AAZ25693  
ID AAT25693 standard; cDNA to mRNA; 226 BP.  
XX AC  
XX AAT25693;  
XX AC  
DT 10-OCT-1996 (first entry)  
XX DT  
DE Human gene signature HUMGS07893.  
XX DE  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
XX KW  
OS Homo sapiens.  
XX OS  
XX WO9514772-A1.  
PN WO9514772-A1.  
XX PN  
PD 01-JUN-1995.  
XX PD  
XX PF 11-NOV-1994; 94WO-JP01916.  
XX PF  
XX PR 12-NOV-1993; 93JP-0355504.  
XX PR  
PA (MATS/) MATSUBARA K.  
PA (OKUBO/) OKUBO K.  
XX PA  
XX PI Matsubara K, Okubo K;  
XX PI  
XX WPI; 1995-206931/27.  
XX WPI  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX PT  
XX PS Claim 1; Page 1908; 2245pp; Japanese.  
XX PS  
XX A single-stranded DNA (or its complementary strand or the corresp.  
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX given in AAT19001-T26837 and which is able to hybridise to part of  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
XX sequences were obtained from 3'-directed cDNA libraries prepared  
XX from various human tissues; synthesis of cDNA was initiated from the  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX untranslated sequence is unique to a particular mRNA species, almost  
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
XX is constructed so as to reflect accurately the relative abundance of  
XX different mRNAs in the particular tissue from which it was derived.  
XX The appearance frequency of a given GS in a cDNA library can be  
XX determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

XX Sequence 226 BP; 43 A; 65 C; 61 G; 56 T; 1 other;

Query Match 2.4%; Score 206.2; DB 16; Length 226;  
Best Local Similarity 97.8%; Pred. No. 1.2e-26;  
Matches 219; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 8326 gatctctgtaccgttcaagaggcgtgtgcagggccagcagtcgtagcccatcactcg 8385  
|||||  
Db 1 gatctctgtaccgttcaagaggcgtgtgcagggccagcagtcgtagcccatcactcg 60  
|||||

QY 8386 caggaccaagggcgaggactgctcg-tcacgcccgcgtgtgctcctccctccctt 8444  
|||||  
Db 61 caggaccaagggcgaggactgctcg-tcacgcccgcgtgtgctcctccctccctt 120  
|||||

QY 8445 ccttgggcagaatgaattcgcgtattctgtggccgcatttgcaggggtgtggtga 8504  
|||||  
Db 121 ccttgggcagaatgaattcgcgtattctgtggccgcatttgcaggggtgtggtga 180  
|||||

QY 8505 ttctgtcatttacacacgctcgttctaataaaaaagcgaattata 8548  
|||||

Db 181 ttctgtcatttacacacgctcgttctaataaaaaagcgaattata 224  
|||||

## RESULT 11

AAZ16912

ID AAZ16912 standard; cDNA; 789 BP.

XX AC

XX AAZ16912;

XX DT 12-OCT-1999 (first entry)

XX DE Human gene expression product cDNA sequence SEQ ID NO:4382.

XX DE

XX KW Human; gene; gene expression product; diagnosis; therapy; probe;

XX KW detection; mapping; tissue typing; profiling; forensic; cancer;

XX KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX OS Homo sapiens.

XX OS

XX PN WO9938972-A2.

XX PN

XX PD 05-AUG-1999.

XX PD

XX PF 28-JAN-1999; 99WO-US01619.

XX PF

XX PR 03-APR-1998; 98US-0080666.

XX PR

XX PR 28-JAN-1998; 98US-0072910.

XX PR

XX PR 24-FEB-1998; 98US-0075954.

XX PR

XX PR 31-MAR-1998; 98US-0080114.

XX PR

XX PA 03-APR-1998; 98US-0080515.

XX PA

XX PA (CHIR) CHIRON CORP.

XX PA

XX PA (HYSE-) HYSEQ INC.

XX PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

XX PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;

XX PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

XX PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

XX PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX DR Novel human genes and their expression products which are

XX PT differentially expressed in different cell types

XX PS Claim 1; Page 2075; 2479pp; English.

XX CC The present invention describes a library of human polynucleotides

XX CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is

CC







## RESULT 13

AAT31931  
ID AAT31931 standard; DNA; 1922 BP.

XX AC AAT31931;

DT 27-SEP-1996 (first entry)

XX DE Retinoid X receptor interacting protein R1P13 DNA.

XX KW Retinoid X receptor interacting protein; RXR; R1P13; ss.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
XX CDS 2..1861  
XX FT /\*tag= a

XX PN W09621677-A1.

XX PD 18-JUL-1996.

XX PF 08-DEC-1995; 95WO-US16311.

XX PR 13-JAN-1995; 95US-0372652.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Choi H, Moore D, Seol W;

XX DR WPI; 1996-342241/34.

XX DR P-PSDB; AAR99738.

XX PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to  
XX PT modulate or mediate RXR function, anti-RIP antibodies can be used to  
XX PT determine RIP subcellular distribution patterns

XX PS Claim 8; Page 56-57; 90pp; English.

XX CC A cDNA clone (AAT31931) codes for retinoid X receptor interacting  
XX CC protein R1P13 (AAR99738), which is a candidate transcriptional  
XX CC co-activator. It was isolated from a mouse liver cDNA library.  
XX CC The cDNA can be used for the recombinant prodn. of R1P13 in  
XX CC transformed host cells.

XX SQ Sequence 1922 BP; 545 A; 508 C; 451 G; 418 T; 0 other;

Query Match 1.7%; Score 142; DB 17; Length 1922;  
Best Local Similarity 52.5%; Pred. No. 1.4e-15;  
Matches 435; Conservative 0; Mismatches 375; Indels 18; Gaps 5;

QY 6730 gctgctgtaccggatgggaacagacgagccagcaggatggctccaaagtctccagg 6789

DB 1039 gctctgtcacagagggagggagccctgtgagcaaggagtgattctcgatcacagg 1098

QY 6790 caacacagccagcccgacctcttcagcaagctacccagagacaactccgcctgg 6849

DB 1099 aagtataagctactgctctcattcttcaccaagct---tgaaagccatcacccctgg 1155

QY 6850 caagtcacaagaagcaagatcaacaagaagctgaacacccacacccggaatgagcctga 6909

DB 1156 taatatcaagaacaagaattttctgaagtgaactcttctgtgaggtgactctga 1215

QY 6910 atacaatcacgcagcctgggagagatcttcaatgctcccccattaccgggaacagg 6969

DB 1216 tatggcgtgctcagccaggaacagagatcttcaatctgcagcagcttaccacatcagg 1275

QY 6970 ccttatgacctatagaagccagcggtgcaggaaacatgccagacacacatggggctgga 7029

DB 1276 tgcagtgactcaagagccattcttctgtgatcccgccag---taacctgtgtctaga 1332

QY 7030 ggcataatagaaggcactcatcgtggtaaat---atgaccagtggaagagtcctccgcc 7086

DB 1333 agacatcatcagaaaggctctcatggtggaagtttgatgataaagttgaagatcatggtgt 1392

QY 7087 gctcagcgccaatgcttttaaccctctgaatgccagtgccagcctgcccgtctgtatgcc 7146

DB 1393 tgtcatgtcccatcctgtgggcattatgctggtagtgccagcactcagtggtgacgag 1452

QY 7147 cataacccgtctgacgagcagagagtgaccacacactcaccctgcgcagtggtgcggcgga 7206

DB 1453 cagcgaggcacggagagatgaaggaggccatcactcatcgaggagtatgcaaacccaaa 1512

QY 7207 ggcacaaggtctctggcagaccagcagccgaaagtcctcccgcccggtggcctggc 7266

DB 1513 gctgatcaacaatacaaacagcaggaagtctaaatctctattctctgggcaagctattt 1572

QY 7267 atctggggaccggccacccctctgtctcagtgctcagtcagtgaggagactgcaaccgcg 7326

DB 1573 aggaactgaaaggccttctctctctgtgctcattcaggaagtgattaccacagga 1632

QY 7327 gacgccgctcaccacccgctgtgggagacagccctcgcgcaggttccacgccatt 7386

DB 1633 gacac-----caggatgggcatgggaagatcgccctcttcaacaggttctactcagtt 1686

QY 7387 ccctacaacccctctgacatcgctcagcggtgctcagcggtgtcatggcttcccccaccacc 7446

DB 1687 ccctacaacccctctgacatcggtgctcagcagtagtacaccacccatcacagatcgcatg 1746

QY 7447 gggctctcccgcggggcagc---gggcccctcgtcgccccaccacacgcctgggagcagga 7503

DB 1747 cgccccatctgcatcaccacagcagctccacatcaacagacgcctctgggagagga 1806

QY 7504 gcccaagccactgctgctgcgagtagacacactctccacagcga 7551

DB 1807 gcttgcctcctctctcagcgcgtagtagacacactgctctgtagtagcga 1854

## RESULT 14

AAF67531

ID AAF67531 standard; cDNA; 342 BP.

XX AC AAF67531;

DT 09-APR-2001 (first entry)

XX DE Novel human polynucleotide, SEQ ID NO: 3293.

XX KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX KW breast cancer; lung cancer; cancer detection; ss.

XX OS Homo sapiens.

XX PN W0200102568-A2.

XX PD 11-JAN-2001.

XX PF 30-JUN-2000; 2000WO-US18374.

XX PR 02-JUL-1999; 99US-0142310.

XX PR 02-JUL-1999; 99US-0142311.

XX PA (CHIR) CHIRON CORP.

XX PA (HYSE-) HYSEQ INC.

XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

XX PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

XX PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

XX PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX DR WPI; 2001-091805/10.

XX PT Library of polynucleotides for diagnosing a cancerous state of a  
XX PT mammalian cell and detecting cancer, particularly of the colon or  
XX PT prostate, comprises 3351 human polynucleotide sequences -



XX	WO9913886-AL.	
PN	XX	
XX	XX	
PD	25-MAR-1999.	
XX	XX	
PF	17-SEP-1998;	98WO-US19419.
XX	XX	
PR	09-JUN-1998;	98US-0093972.
PR	17-SEP-1997;	97US-0059160.
XX	XX	
PA	(UYEC-) UNIV EAST CAROLINA.	
XX	XX	
PI	Nyce JW;	
XX	XX	
DR	WPI; 1999-229400/19.	
XX	XX	
PT	New antisense oligonucleotides used in treatment of, e.g. pulmonary	
PT	vasoconstriction	
XX	XX	
PS	Disclosure: Page 37; 120pp; English	

The specification describes antisense oligonucleotides (AAX52869-X55271), directed against at least 2 mRNAs selected from target genes, coding and non-coding regions of RNAs corresponding to target genes, gene initiation codons, genomic flanking regions, intron-exon borders, the

3'-end, the 3'-end and the juxta-section between coding and non-coding regions and all segments of RNAs encoding proteins associated with one or more diseases, conditions or mixtures. The antisense oligonucleotides may be derived from sequences AAX5272-74. These multiple target oligonucleotides (specifically AAX55180-271) can be used for the antisense treatment of diseases and conditions. Typical diseases and conditions are those associated with impaired respiration and inflammation, including lung diseases, pulmonary vasoconstriction, allergic rhinitis, acute asthma, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as well as all types of cancers which may metastasize or have metastasized to the lungs, including breast and prostate cancer.

xx Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;

SQ

Query Match	1.4%;	Score 118.4;	DB 20;	Length 114955;
Best Local Similarity	29.7%;	Pred. No. 3.1e-11;		
Matches 972;	Conservative 280;	Mismatches 1987;	Indels 332;	Gaps

Db 111316 GCTGCCACCGCCCCCAAGTGGAGTGGCTGGGGGCCACVCVAAACGGGACCG 111257

Qv 2403 aagccacggagagccctagcccccacacacacccccatccccctctccacctcctcctg 2452

[illegible]

Db 111196 NNNSCCGACGGCCCGGAGCGGACGGCCNNHNNNSCGCGCGCGGGAGAGACNN 111137

[illegible]

DB LIIU16 CNNHNNNSCCGGCAGCGGCGGCGGCCGCGCAAGAGGGGCCCGACCGCG I10957



QY 2703 gacggccaccactgccaagagctggcgccccccacagacagcactcagtgctaact 2762  
Db 110956 GCAGCCNNHHNNNSCCCGCGCGCGCGCGCGCGCGGAGNNHHNNNSTCGCGGAGCGCGC 110997  
QY 2763 gcaagtgcagacagagtgatagagcgg--agggcggcagacaagaacccggctgtctcccc 2820  
Db 110896 GCGCAGCAGCCNNHHNNNSGGCGCGCGCGGAGGCACGCCACCCGCCNNHHNNNSCCCGC 110837  
QY 2821 aagggccagctctcaccgcagactggcgacccccggcgcccaatgcctcaccgccagaacc 2880  
Db 110836 GCAGVGCG 110777  
QY 2881 actggacactgaagcagctgaagcagcagcagcagcagcagcagcagcagcagcagcagc 2940  
Db 110776 GCGAGCCGACVGGGGCGACG 110717  
QY 2941 agtccatagccccccgggagagcagcagcagcagcagcagcagcagcagcagcagcagc 3000  
Db 110716 GCGCCNNHHNNNSCCCG 110657  
QY 3001 gccaccgcaaacctgcagcgcagagagcagcagcagcagcagcagcagcagcagcagcagc 3060  
Db 110656 GCGCGCGNNHHNNNSGGCG 110597  
QY 3061 gggcaagagcagggagcccgccagcccgccgcagacagagcagcagcagcagcagcagcagc 3120  
Db 110596 GCGCGCGNNHHNNNSGGCG 110537  
QY 3121 gaagctgctgggaacccctgtgtgacttcggcctgcgcttcgagcagcagcagcagcagc 3180  
Db 110536 GCGGCG 110477  
QY 3181 tgagtgatcaagcctcccgcatgcccgcagccctcagcctcctcctcagcctcctcagc 3240  
Db 110476 GGGCGCCNNHHNNNSGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 110417  
QY 3241 tggtaaccaatgcccctggcctcctcagcagcagcagcagcagcagcagcagcagcagcagc 3300  
Db 110416 GGGCGCCNNHHNNNSGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 110366  
QY 3301 caccatctccaaacgcctccctcctcctcctcctcctcctcctcctcctcctcctcctc 3360  
Db 110365 CNHHNNNSGGCG 110306  
QY 3361 gcaaatagtgctcctcccaaggaatgcttcagcagcagcagcagcagcagcagcagcagcagc 3420  
Db 110305 GCG 110246  
QY 3421 tgcgaaggcccggtgggcccctgtcaccatggggtgcccctgcccctgcccctgcccctgccc 3480  
Db 110245 NNHHNNNSGCCACGAGNNHHNNNSGCCACGAGCAGCAGCAGCAGCAGCAGCAGCAGC 110186  
QY 3481 gctggcacccttcagcagagtgagcagcagcagcagcagcagcagcagcagcagcagcagc 3540  
Db 110185 NNNSGCCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 110126  
QY 3541 accggagagctgggggtgcacacagccagcagcagcagcagcagcagcagcagcagcagcagc 3600  
Db 110125 CVNNHHNNNSGCCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 110066  
QY 3601 gggctcagttccgggggaagcacaacaaaggcattccacagcacacagcagcagcagcagcagc 3660  
Db 110065 CCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 110006  
QY 3661 cagcgcatacacaaccggctcctacacacagcagcagcagcagcagcagcagcagcagcagcagc 3719  
Db 110005 GAGCVCVVGCGCGNNHHNNNSGCCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109946  
QY 3720 ---agggcaccatcaccagagatcatcgccgagcagcagcagcagcagcagcagcagcagcagc 3776  
Db 109945 CVCVVGCGCGACNNHHNNNSCVCVVGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 109886  
QY 3777 gggagagcagcgtcccaaggggcagcgtcatctcagaaaggcagagggcagcagcagcagcagc 3836

Db 109885 VCGCGCGACNNHHNNNSGVCVVGCGCGCGACNNHHNNNSAGCVCVVGCGCGCGACNNHHNN 109826  
QY 3837 cctatgaaggggtgagtgctgtgacccagtgctcacaaggagacagcagcagcagcagcagcag 3896  
Db 109825 AGCVCVVGCGCGACNNHHNNNSGAGCVCVVGCGCGCGACNNHHNNNSAGCAGCVCVVGCG 109766  
QY 3897 gaccccccatagacgagcgccgcccccaagcagcactatgacatgatgagagcagcagcagcag 3956  
Db 109765 GACNNHHNNNSAGCAGCVCVVGCGCGCGACNNHHNNNSGAGCAGCVCVVGCGCGCGAC 109706  
QY 3957 gagagcactctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4016  
Db 109705 NNNSAGCAGCAGCVCVVGCGCGCGACNNHHNNNSAGCAGCAGCVCVVGCGCGCGACNN 109646  
QY 4017 gacacagcccccaacacactcaaaagcagcagcagcagcagcagcagcagcagcagcagcagc 4076  
Db 109645 SCCAGCAGCAGCVCVVGCGCGCGACNNHHNNNSGAGCAGCAGCVCVVGCGCGCGACNN 109586  
QY 4077 tccctcggtcctacgtgagggcacagagagcagcagcagcagcagcagcagcagcagcagcagc 4136  
Db 109585 NSGGCGCGCGVCGCCACCCNNHHNNNSGGCGCGCGCAGNNHHNNNSGGCGCGCGCAGNN 109526  
QY 4137 agcgaggagggcacgctccgccccccacgcctcagcgagcagcagcagcagcagcagcagcagc 4196  
Db 109525 SGGGCGCCVAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109466  
QY 4197 qcagggcctgggccccctgaagctgaagcagcagcagcagcagcagcagcagcagcagcagcag 4256  
Db 109465 SGGGCGCCVAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109406  
QY 4257 -----aggagcgggcgccctcctcctcagcagcagcagcagcagcagcagcagcagcagc 4304  
Db 109405 GCCNNHHNNNSGGGCGCCVAGGCGACGCCANNHHNNNSGGGCGCCVAGGCGAGCAGCAG 109346  
QY 4305 aacgcccagagtgccccggccccggcgctcagagggggtccatcacgcaggggca 4364  
Db 109345 SGGGCGCCVAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109286  
QY 4365 ccccgctcaagtaagacacccggcggtccacactggctccaaaagcagcagcagcagcagcagc 4424  
Db 109285 CCCVAGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109226  
QY 4425 cctcatcgagcagccccggcgagcgttccacacgcgtgcacccgctgagatgagtgagcag 4484  
Db 109225 CCVAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109166  
QY 4485 acgcccggcactggaaacgtgctgtctcagagagagcagcagcagcagcagcagcagcagcagc 4544  
Db 109165 GCGCCVAGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 109106  
QY 4545 ccagcagctcggggggtccattgctgagcgcgccccgggtcattgtgctgagctgagtgta 4604  
Db 109105 HNNNSGGGCGCCVAGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 109046  
QY 4605 agcgcgagcagcagccccctgacattagagcagcagcagcagcagcagcagcagcagcagcagc 4664  
Db 109045 CGCGCGCGNNHHNNNSGGGCGCCVAGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 108986  
QY 4665 cagaggttcgcccgtgaccatgaggggagcccaagcccgctgagcagcagcagcagcagcagc 4724  
Db 108985 VAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 108926  
QY 4725 cgtccagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4784  
Db 108925 CCCNNHHNNNSGGGCGCCVAGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 108866  
QY 4785 cccgcagc 4844  
Db 108865 GCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 108806  
QY 4845 ttcggggcggtgagtgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4904



Db 108805 CCCCCNNNNNNNSCGCCGCCGCCNNNNNNNSGCCGCCGCCNNNNNNNSCGCCGCCGCC 108746  
Qy 4905 ctcatacccgcgccatccctctgagcagcgctgtactactacgtgccccgacacc 4964  
Db 108745 GCNNNNNNNSCGCCG---CCCCGCCNNNNNNNSCGCCGCCGCCNNNNNNNSCGC 108690  
Qy 4965 tggcccccaacccccaccctaccgcaacctgtaccaccctacatcatcgggctaccocg 5024  
Db 108689 CGCCGCCGCCGCCNNNNNSGCGCGCCGCCGCCNNNNNSGCGCGCGCGCGCGCGCG 108630  
Qy 5025 acacggcggtgagagacggagacacatcatcaatactacatcaacctcgacagaga 5084  
Db 108629 CCNNNNNSAGCGCGCGCGCGGCCGCCNNNNNSAGCGCGCGCGCGGCCGCCNNNN 108570  
Qy 5085 tgcaccacaacgcccacgcgcacatggcccagcgagctgatatgtgagggcctctgc 5144  
Db 108569 NSCCAGGCGCGCGCGCGGCCGCCNNNNNSGCGCGCGCGCGCGCGCGGCCGCCNNNNNSA 108510  
Qy 5145 cccggagtcctgtggaactcaactacgtgcgggtcccgagggatcatcgcacctgt 5204  
Db 108509 GCCAGGCGCGCGCGCGGCCGCCNNNNNSGAGCCAGGCGCGCGCGCGCGCGGCCGCCNNNNNS 108450  
Qy 5205 cccaagtccacacctgtctgtctgtgccccgacaccagggacccacgacccgcca 5264  
Db 108449 CGAGCCAGGCGCGCGCGCGCGGCCGCCNNNNNSGCGA-GCCAGGCGCGCGCGCGCGGCCCN 108391  
Qy 5265 tggacgccttgcctacctccccacgcgcgcgcaccttcagcagcgccacagcagct 5324  
Db 108390 NNNNSGGGAGCCAGGCGCGCGCGGCCGCCNNNNNSAGCGCGCGCGCGCGCGCG 108331  
Qy 5325 cccactctccccaggggtccacacacttgacaaaaaacaccaccgctctcgtccg 5384  
Db 108330 CCCCCGCCNNNNNSVAGCGGAGCCAGCGCGCGCGCGCGCGGCCGCCNNNNNSVAGCGGAG 108271  
Qy 5385 agcgggagcagacgggatcgagagcgggaccgggagtcgggagcgggaaaaagtccatcc 5444  
Db 108270 CCAGCGCGCGCGCGCGCGGCCGCCNNNNNSCCVAGGCGAGCCAGGCGCGCGCGCGCGCN 108211  
Qy 5445 tcacgtccaccacgacgtggagcagcaccatctggagacctgtgtacagagcagaga 5504  
Db 108210 NNNNSCCCVAGCGGAGCCAGCGCGCGCGCGCGCGGCCGCCNNNNNSGCCCVAGGCGAGCCA 108151  
Qy 5505 gcggcagcagcgagcagcgcggggtgggggagcagcagcggcccgccctccact 5564  
Db 108150 GCGCGCGCGCGCGCGGCCGCCNNNNNSGGCCVAGGCGAGCCAGGCGCGCGCGCGCGCN 108091  
Qy 5565 cccatgccaccacgacctgcctctccc 5595  
Db 108090 NNNNSGGGCCCVAGGCGAGCCAGGCGCGCG 108060

Search completed: September 8, 2001, 09:57:07  
Job time: 1139 sec







Db 1039 GCTCTGTGCACAGAGGGAGTGGACCTCTGTGAGCAAAAGAGTGAFTCTCGATCACCAGG 1098  
QY 6790 caacaccagccagccagcctctctcagcaagctgacagagcaacacccgcatgct 6849  
Db 1099 AGTATAAGCTACTTCCCTCACTTCTACCAAGCT---TGAAGCAGCATCACCATGCT 1155  
QY 6850 caagtccaagaagaagatcaacaagaagctgaacaccccaacccggaatgagcctga 6909  
Db 1156 TAAATCAAGAAACAGAGAAATTTTCGTAAGTGAAGTGAAGATCATGCTCTGA 1215  
QY 6910 atacaataacagccagcctggaacagagatcttcaatatgcccgcacacccggaacag 6969  
Db 1216 TATGGCAGCTGTCTGAGCAGGAGAGATCTTCAATCTGCCAGCAGTATACCATCAGG 1275  
QY 6970 ccttatgacctataaagcagcagcagcagcagcagcagcagcagcagcagcagcag 7029  
Db 1276 TGCAGTGAAGCTCAAGAGCCATCTTTTCTGATCCCGCAG---TAACTTGGTCTAGA 1332  
QY 7030 ggcataatagaagcactcatggtgtaaat---atgaccagtggaagagtcgccgccc 7086  
Db 1333 AGACATCATCAGAGAGGCTCTCATGGAGTGTGATGAAGTGAAGATCATGCTGT 1392  
QY 7087 gctcagcgcgaatgcttttaacccctgtaagcagctgacagcagcagcagcagcag 7146  
Db 1393 TGTCTATGCCATCTCTGTGGCATTATGCTGTGATGCGCAGCAGCAGTGTGACGAG 1452  
QY 7147 cataaccgctgctgacgagcagcagcagcagcagcagcagcagcagcagcagcag 7206  
Db 1453 CAGCGAGGACAGGAGAGATGAAGGGAGGAGCCATCACCCTCATGAGGAGTATGCAAAACCAA 1512  
QY 7207 ggcgaagctctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 7266  
Db 1513 GCTGATCAACAATCAACAGCAGGAGTCTAAATCTCTATCTCTGGGCAAGCTATTT 1572  
QY 7267 atctggggaccgcccacccctctctcagtcagcagcagcagcagcagcagcagcag 7326  
Db 1573 AGGAAGTGAAGGCT 1632  
QY 7327 gacgcgctcaccacccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7386  
Db 1633 GACAC-----CAGATGGGATGGGAAGATCGGCCCTCTTCAACAGGTTCTACTCAGT 1686  
QY 7387 cccctacaacccctgactgctgctgctgctgctgctgctgctgctgctgctgctgctg 7446  
Db 1687 CCTTACACCTCTGACCATACGAGTGTCTACAGCTACACCATCAGCATGCGATG 1746  
QY 7447 gggctcccccggcagc---ggcgcctctgctgctgctgctgctgctgctgctgctg 7503  
Db 1747 CCGCCCATCTGCATCACCACCAAGCAGCTCCACATCAACAGAACCCGATCTGGGAGGGA 1806  
QY 7504 gcccagccactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 7551  
Db 1807 GCCTGCCCGCTCTCTCAGCCAGTATGAGACACTGTCTGATAGGA 1854

## RESULT 2

PCT-US95-16311-9

; Sequence 9, Application PC/TUS9516311

; GENERAL INFORMATION:

; APPLICANT: Moore, David

; APPLICANT: Seol, Wongi

; APPLICANT: Choi, Hwang-sik

; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING

; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street, Suite 3100

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16311  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1922 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

PCT-US95-16311-9

Query Match

Best Local Similarity 1.7%; Score 142; DB 5; Length 1922;

Matches 435; Conservative 0; Mismatches 375; Indels 18; Gaps 5;

QY 6730 gctgctgtaccgggagtggaacagacagagccagcagagatgggctccaaagtctccagg 6789

Db 1039 GCTCTGTGCACAGAGGGAGTGGACCTCTGTGAGCAAAAGAGTGAFTCTCGATCACCAGG 1098

QY 6790 caacaccagccagccagcctctcagcaagctgacagagcaactccgcatgct 6849

Db 1099 AGTATAAGCTACTTCCCTCACTTCTTCCACCAAGCT---TGAAGCAGCATCACCATGCT 1155

QY 6850 caagtccaagaagaagatcaacaagaagctgaacaccccaacccggaatgagcctga 6909

Db 1156 TAAATCAAGAAACAGGAAATTTTCGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1215

QY 6910 atacaatatcaccagcctggagcagagatcttcaatatgcccgcacacccaggaacag 6969

Db 1216 TATGGCAGCTGTCTGAGCAGGAGAGATCTCAATCTGCCAGCAGTATACCATCAGG 1275

QY 6970 ccttatgacctatagaagccagcagcagcagcagcagcagcagcagcagcagcagcag 7029

Db 1276 TGCAGTGAAGCTCAAGAGCCATCTTTTCTGATCCCGCAG---TAACTTGGTCTAGA 1332

QY 7030 ggcataatagaagcactcatggtgtaaat---atgaccagtggaagagtcgccgccc 7086

Db 1333 AGACATCATCAAGAGGCTCTCATGGAAAGTGTGATGATGAAGTGAAGTGAAGTGAAGTGA 1392

QY 7087 gctcagcgcgaatgcttttaacccctgtaagcagctgacagcagcagcagcagcagcag 7146

Db 1393 TGTCTATGCCATCTCTGTGGCATTATGCTGTGATGCGCAGCAGCAGTGTGACGAG 1452

QY 7147 cataaccgctgctgacgagcagcagcagcagcagcagcagcagcagcagcagcagcag 7206

Db 1453 CAGCGAGGACAGGAGATGAAGGGAGGAGCCATCACCCTCATGAGGAGTATGCAAAACCAA 1512

QY 7207 ggcgaagctctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 7266

Db 1513 GCTGATCAACAATCAACAGCAGGAGTCTAAATCTCTATCTCTGGGCAAGCTATTT 1572

QY 7267 atctggggaccgcccacccctctctcagtcagcagcagcagcagcagcagcagcag 7326

Db 1573 AGGAAGTGAAGGCT 1632

QY 7327 gacgcgctcaccacccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7386















Db 19697 AGGAGTTAGAGGAGGAGGAGGAGGAGGAGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGG 19638  
Qy 1239 acggcttatgccaccccatgaagtgtaacaaagacgcgcagtcacatgaatgtga 1298  
Db 19637 ACCAGGTTAGAGGAGGAGGAGGAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGC 19578  
Qy 1299 gtgagcagagagagagacccctccggggaagattcatgcagcatcccaagaactttg--- 1355  
Db 19577 AGGAGCAGGAGTTAGAGGAGGAGGAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGG 19518  
Qy 1356 ----gctatgcacatctctctggagagaaagacagtgctgagtcctctcttact 1412  
Db 19517 AGCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGC 19458  
Qy 1413 acctactaagaagaatagaaactataagacctgtgtgagacgagctatcgccgcgcg 1472  
Db 19457 AGAGCAGGAGTTAGAGGAGGAGGAGGAGGAGGAGGAGTTAGAGGAGGAGTTAGAGGAGGTTGG 19398  
Qy 1473 gcaagagccagcagcaacacagcagcagcagcagcagcagcagcagcagcagcagcagc 1532  
Db 19397 AGGAGCAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTAGAGGAGTTGGAGAGC 19338  
Qy 1533 ccatgccccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1592  
Db 19337 AGGAAGAGCAGGAGTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTAGAGGAGGTTGG 19278  
Qy 1593 aggagagagagagagc 1652  
Db 19277 AGAGCAGGAGCAGGAGTTAGAGGAGGAGTTAGAGGAGGAGTTAGAGGAGGAGTTAGAGGAGGTTGGAGC 19218  
Qy 1653 a 1653  
Db 19217 A 19217

## RESULT 6

US-09-050-863-2  
; Sequence 2, Application US/09050863  
; Patent No. 6114111  
; GENERAL INFORMATION:  
; APPLICANT: Lao, Ying  
; APPLICANT: Hiang, Betty  
; APPLICANT: Payan, Don  
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
; TITLE OF INVENTION: System  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/050, 863  
; FILING DATE: 30-MAR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 949-8711  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
US-09-050-863-2

Query Match 1.1%; Score 96.2; DB 3; Length 2580;  
Best Local Similarity 48.2%; Pred. No. 5.6e-09;  
Matches 331; Conservative 0; Mismatches 333; Indels 3; Gaps 2;

Qy 1476 agagccagcagcaacacacagcagcagcagcagcagcagcagcagcagcagcagcagccca 1535  
Db 662 AGGAGCGGGGGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 721  
Qy 1536 tgcgccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1595  
Db 722 AGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 781  
Qy 1596 aggaggaagccggaggtggagaaacagcagcagcagcagcagcagcagcagcagcagcagc 1655  
Db 782 AGGAGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 841  
Qy 1656 acacctcggggagagacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 1715  
Db 842 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 899  
Qy 1716 ccaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1775  
Db 900 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 958  
Qy 1776 gcgagggagccatcaccccccagcagcagcagcagcagcagcagcagcagcagcagcagc 1835  
Db 959 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1018  
Qy 1836 gttctcgtgacagaagaagaataagaaacagcagcagcagcagcagcagcagcagcagcagc 1895  
Db 1019 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1078  
Qy 1896 gcaactggtcgccatcccgccgagtggtgggtcccaagactgtgtcagctgtaagaact 1955  
Db 1079 AGGAGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1138  
Qy 1956 tctacttcaactacaagaagagggcagaaacctcgatgatcttgcagcagcagcagcagcagc 2015  
Db 1139 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1198  
Qy 2016 agatggagaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2075  
Db 1199 AGGAGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1258  
Qy 2076 aggaggtgcatcccgcccggtgtgagagatgagagatgagagcgtcggtgagcgtgagcgt 2135  
Db 1259 AGGAGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1318  
Qy 2136 gaaatgagagagagatggtggagagag 2162  
Db 1319 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1345

## RESULT 7

US-09-130-114-1/c  
; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Damaj, Bassam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; TITLE OF INVENTION: From Multiple Transfected Episomes  
; FILE REFERENCE: 0867/1D903US1  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36



























GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2001, 09:38:08 ; Search time 5930.11 Seconds  
(without alignments)  
13662.547 Million cell updates/sec

Title: US-09-522-753-4  
Perfect score: 8571  
Sequence: 1 catgtcgggtccacacagc.....caaaaaaaaaaaaaaaaa 8571

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	gb_estl1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est33:*
26:	gb_est34:*
27:	gb_est35:*
28:	gb_est36:*
29:	gb_est37:*
30:	gb_est38:*
31:	gb_est39:*
32:	gb_est40:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*
44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estom1:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.











Qy	7184	aactcgccaagtgtgcgctggcgaagccaaagggtctcttgcagaccacagcagccgaaaaagccc	7243
Db	544	ACCTCGCAGGTGGCGGCGGAAGCAAGGTCCTTGGCAGACCAGCAGCCGAAAGGCC	603
Qy	7244	aagtccccggccccggccttgcatctgggaccggccacctctgtctctcaatgcac	7303
Db	604	AAGTCCCCGGCCCCGGCCTTGGCATCTGGGACCGG-CACCCCTCTGTCTCTCTCAATGCAC	662
Qy	7304	tcgaggaggagactgcaaccgcgcggagcgcgcctcaccaaacccgcgtgtgggagagacaggccc	7363
Db	663	TGGAGGAGACTGTCAACCGC--GGACGCGTGTCACCAACCGCGTGTGGGAGGACAGGCC	720
Qy	7364	tcgtccgcagttccacgcattcccctcacaaacccccctgatcatcggtcgcaggcgggt	7423
Db	721	TCGTCCGCGAC--TTACAGCCATTTCCTTAATCCCTGATCATCGGTGCGGCGTGCAGGG	778
Qy	7424	gtc 7426	
Db	779	TTC 781	
RESULT	4		
LOCUS	BE728145	804 bp mRNA EST	15-SEP-2000
DEFINITION	601563413f1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3833162 5'		
ACCESSION	BE728145	mRNA sequence.	
VERSION	BE728145		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomom Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1. (bases 1 to 804)		
AUTHORS	NIH-MGC http://mhc.nci.nih.gov/		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-re@mail.nih.gov Tissue Procurement: ATCC/DCTD/DTTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LLCW511 row: i column: 03 High quality sequence stop: 716.		
FEATURES	Location/Qualifiers		
source	1..804		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3833162"		
	/clone_lib="NIH_MGC_20"		
	/tissue_type="melanotic melanoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: skin; Vector: pOTEB7; Site_1: XhoI; Site: ECORI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong i the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)		
BASE COUNT	200 a 236 c 262 g 106 t		
ORIGIN			
Query Match	7.8%; Score 668.4; DB 139; Length 804;		
Best Local Similarity	92.6%; Pred No. 3.2e-115;		
Matches 746; Conservative	0; Mismatches 56; Indels 4; Gaps		
Qy	242	gagctccacatcgcgccagagtcctccactcatactgccagctgggaagtcagagatg 301	











Query Match	7.5%	Score 645;	DB 150;	Length 746;
Best Local Similarity	95.4%;	Pred. No. 7.6e-111;		
Matches 707;	Conservative 0;	Mismatches 30;	Indels 4;	Gaps 4;
QY 7829	ggaggggggcaagcgagcggtgtggcagcacacaca-ggcgcgccagcgcgccagggac	7887		
Db				
6	CGTGTGGGGCAGCGAGCGGTGTGGCAGCCACACGACGCGGGCGGACGCGCCAGTGAC	65		
QY 7888	ccaaagcaggatgaccacgcacctccacgcccactgcctccccccgaatgcatcttggaaacca	7947		
Db				
66	CCAAAGCAGGATGACACGACACCTCCACGCCACTGCCTCCCCCGAATGCATTTGGGAACCA	125		
QY 7948	aagtcctaaactgaactcgacgcccccgcgccctccctccgctcccatccagccttagcgc	8007		
Db				
126	AAGTCTAAACTGAGCTCGACGCCCGCGGCCCTCCCTCCGGCTCCATCCCGCTTAGCGC	185		
QY 8008	cttggaacagatggscgagggccctgtccagcccccaagtcgcgtcgtccgttccccacag	8067		
Db				
186	CTGGACAGATGGACGCGAGGCCCTGTCCAGCCCCCAGTGCCTCGTCTGTGTCGCCACAG	245		
QY 8068	actgccccagccaaacgagattgctggaaaaccaagtgcaggcccaggtggcgcgacaaaaagg	8127		
Db				
246	ACTGCCCCAGCCAAACGAGATTGCTGTGAAACCAAGTCAAGCCAGGTGGCGGCGACAAAAGGG	305		
QY 8128	ccaggtgcgccttgggggaacgagatctccgagagactggagactttttttcacacatcg	8187		
Db				
306	CCAGGTGCGGCTGGGGGAACGAGTGCCTCGAGGACTGGAGTGTGTTTTTTCACACATCG	365		
QY 8188	ttgcgcgagcgtggggaagaaagcgagatgtaaatgatgtgttggtttacaggggtatat	8247		
Db				
366	TTGCCGCGAGCGTGGGAAGGAAGGACAGATCTAAATGATGTGTCGGTTTACAGGGTATAT	425		
QY 8248	ttttgatacctcaatgaattaatctagatgttttaagcaaggaa-ggaacttaccagta	8306		
Db				
426	ATTGTGATACCTTCAATGAATTACTTTCAGATGTTTTTCGCAAGGAAGCGACTTATCCACGTA	485		
QY 8307	ttactgctgtgcttttgatctctgtaccgttcaagagcgctgtgcagggccgacag	8366		
Db				
486	TTACTGTGTGTGCTTTGGTTCCTGCTGCTTACCCTTCAGAGAGGCGTGTGACAGCGCGACAG	545		
QY 8367	tcggtgaccccatcaactcgagggaccaagggcgggagctgctcg-tcaagccccgcgtg	8425		
Db				
546	TCGGTGAACCCATCACCTCGACGAGCAAGGGGGCGGAGCTGCTGGCTTCACGCCCGCGTG	605		
QY 8426	tgctccctccctcccttcttgggcagaatgaaatcgatcgcttatctctgtggccgcca	8485		
Db				
606	TGTCCCTCCCTCCCTCCCTTCCCTTGGGCGAGATGAATACGATGCGCTATCTGTGGCGCGCAT	665		
QY 8486	tttgcgaggtgtgtgttctctgtatttacc-acacgtgcttcttaataaaagccaat	8544		
Db				
666	CTTGGCGCAGGGTGGTGTATTCTGCGATTTTACAACAGTCGCTTCTTAATTACAGAGCGAAT	725		
QY 8545	tatactcaaaaaaaaaaaaaa	8565		
Db				
726	TATACTCCAAGAACAAACAAA	746		



```

Db 301 CTGGCTCCCGCCGAATGCAATTTGGAAACCAAAAGTCTAAACTGAGCTCGACGCCCGCCGCC 360
Qy 7980 tccctcgctcccatcccgcttagcgtctgagacagatggagcagcgccctgtccagcc 8039
Db 361 TCCTCCGCTCCCATCCGCTTAGCGCTCTGGACAGATGGACGAGGCCCTGTCCAGCC 420
Qy 8040 cccagtgcgtcttcctccggtcccccacagactgccccagcccaacagatgctggaaacca 8099
Db 421 CCCAGTGGCGCTGTTCCGCTCCCGCCACAGACTGCCCCAGCAGAGATGCTGTGGAACCA 480
Qy 8100 aqtcaggccagtggtggagacaaaaggccagtgctgcgctgggggaacgagatgctccg 8159
Db 481 AGTCAGGCGAGTGGCGGACAAAAGGGCCAGGTGCGGCTGGGGGAACGAGATGCTCG 540
Qy 8160 aggaactgactgtttttttcacacatcgtgctgcgcagcggtgggaagaaagcagatgt 8219
Db 541 AGGACTGGACTGTTTTTTCACACATCGTTGCGGAGCGGTGGGACGACGCCGATGT 600
Qy 8220 a-aatgatgtgtgtttacagggtatatattttgtataccttcaatgaatgaattcagatg 8278
Db 601 ACCCTGATGTGTGTTACAGGCTATATCTTTGATACCTTCCATGAATTAATTCAGATG 660

RESULT 9
BE391091 688 bp mRNA EST 21-JUL-2000
LOCUS 601286040F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3607929 5',
DEFINITION mRNA sequence.
ACCESSION BE391091
VERSION BE391091.1 GI:9336456
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM259 row: p column: 10
High quality sequence start: 8
High quality sequence stop: 684.
FEATURES
Location/Qualifiers
1..688
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3607929"
/clone.lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 168 a 213 c 208 g 99 t

Query Match 7.4%; Score 630.4; DB 167; Length 688;  
 Best Local Similarity 99.3%; Pred. NO. 4.le-108;  
 Matches 675; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

```

Qy 245 ctccacctggccagagtcacactcatactgcccagctgggaagtcagagatggag 304
Db 6 CTCCACCTGGCGCCAGAGTCCACACTCATACCTGCCGAGCTGGGGAAGTCAGAGATGGAG 65
Qy 305 ttcaatgaaagcaagcgccctcggttagctgctgctgagccccctgctgcaacgtca 364
Db 66 TTCAATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCCCTGACCCCTGCTGGACCGTCA 125
Qy 365 cccctgctggccagggccagcctcggtggtgagacctcaccaagacgtagctg 424
Db 126 CCCCCTGCTGGCCACGGGCCAGCCTCGGGATCTGAAGACCTCACCAAGACCGTAGCCCTG 185
Qy 425 acgggcaagctggaaccgggtgtctccccccagccccccgacactgacctgagctggag 484
Db 186 ACGGGCAGAGCTGGAACCGGTGTCTCCCCCAGCCCGCCACACTGACCTGAGCTGGAG 245
Qy 485 ctggtgcgcacgctgtcccaaggagagctgataccagaacatggaccgctggaccga 544
Db 246 CTGGTCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGACCGCTGGACCGA 305
Qy 545 gagatcaccatgttagagcagcagatctctaagctgaagaagaagcagcaacagctggag 604
Db 306 GAGATCACCATGTTAGAGCAGCAGATCTCTAAGCTGAAGAAGACGACACAGCTGGAG 365
Qy 605 gagggagctgccaagcccgccgagcctgagaaagccgtgtcaccgcccacatcagtcg 664
Db 366 GAGGAGGCTGCCAAGCCCGCAGGCTGAGAAGCCGCTGACCGCCGCCCATCGAGTCG 425
Qy 665 aagcaccgcagcctggtgcagatcatctacgacgagaaacccggaagagctgaaactgca 724
Db 426 AAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCCGGAAGAGCTGAAGCTGCA 485
Qy 725 catcgattctggaaggcctggggcccaagctggagctgcccctgtataaacagccctcc 784
Db 486 CATCGGATTTCTGGAAGGCTTGGGGCCCGCCAGGTGGAGCTGCCCTGTACTAACACAGCCCTCC 545
Qy 785 gacacccgcagctatcatgagaaacataaaacccagcgatcggaagagctaatc 844
Db 546 GACACCCGGCAGTATCATGAGAACATC-AAATTAACAGCGCATGCGGAAGAGTAATC 604
Qy 845 ttgtacttcaagagaggaatcacgcctcggaacaaatggaagcagaagtctgcagcgc 904
Db 605 TTGTACTTCAAGAGGAGG-ATCACGCTCGAAAC-ATGGGAGCAGAG-TCTGCCAGCGC 661
Qy 905 tatgaccagctcatggagcg 924
Db 662 TATGACCAGCTCATGGAGGC 681
```

```

RESULT 10
BE530324/c 657 bp mRNA EST 11-DEC-2000
LOCUS 602071630F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4214551
DEFINITION 5', mRNA sequence.
ACCESSION BE530324
VERSION BE530324.1 GI:11617698
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 657)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
```







```

QY 8060 cccacagagctcccccagcaaacagagattgctgggaaacaaagtcaggcaggtggcgga 8119
Db 417 CCCACAGAGTCCCCAGCAGCAAGAGATTGCTGGAACCAACCAAGTCAGGCCAGGTGGCGGA 476
QY 8120 caaaagggccaggtgctggcctggggggaacagagatgctccgaggaactgactg-ttttttt 8178
Db 477 CAAAAGGGCCAGGTGCGGCTGGGGGGAACGAGATGCTCCGAGACTGGACTGTTTTTTTTT 536
QY 8179 cacacatgctgctccagcagcgggtggaagaaagcagatgt-aaatgatgttgtgt-tt 8236
Db 537 CACACATGCTGCCAGCGGTGGGAAGAAAGGAGATGTAATAATGATGTGTGTCAT 596
QY 8237 acagggtatatttttgataccctcaatgaattaattcag--atgttttacgcaagaa-g 8293
Db 597 CCAGGGTATATTTTGATACCTTCAATGAATTAATTCAGGATGTTTTTACGCAAGGAAG 656
QY 8294 gacttaccagattactgctgctgtgttttgatctctgttaccgttcaag-aggcgt 8352
Db 657 GACTTACCAGTATTACTGGTCTGTGCTGTGCTTGTGATCTCATGTTCATGTTCAAGAAGGCGT 716
QY 8353 gtgcagggccagcagctgctga-ccccatcactcgcagaccagggcgggagctgctc 8411
Db 717 TTGCAGGGCCGACAGTCGGTGACCCCAATCATCGCAGACCAAGGGG-GGGGACTGGTG 775
QY 8412 gtacgccccgctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 8471
Db 776 GCTCAGGCCCGCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8471
QY 8472 tctctg 8476
Db 836 ATCTG 840

```

RESULT 12

```

A1830862
LOCUS wj61b09.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407289 3'
DEFINITION similar to TR:000613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
; mRNA sequence.
ACCESSION A1830862
VERSION A1830862.1 GI:5451533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 622)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the i.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2006 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 465.
Location/Qualifiers
1..622
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407289"
/clone_lib="NCI-CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"

```

FEATURES

source

RESULT 13

BE795997

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

```

BE795997 629 bp mRNA EST 20-SEP-2000
LOCUS 601591169F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945305 5',
DEFINITION mRNA sequence.
ACCESSION BE795997
VERSION BE795997.1 GI:10217195
KEYWORDS EST.
SOURCE human.

```

```

/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 138 a 228 c 171 g 85 t
ORIGIN

```

```

Query Match 7.2% Score 619.4; DB 102; Length 622;
Best local Similarity 99.8%; Pred. No. 4.7e-106;
Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6349 cagccgcgtctccagaccgccccaggggtcaaaaggtcaccagcggtgtgtcacccctgac 6408
Db 1 CAGCCCGCTGCTCCAGACGCCGCCAGGGGTCAAAAGGTCAACAGCGGTGTCACCCCTGGC 60
QY 6409 ccagcacatcagtgagtgatcacacagagactacaccccgccacccacccacagcgtcag 6468
Db 61 CCAGCACATCAGTAGGTGTCATCACAGACTACACCCGGCACACCCACAGCAGCTCAG 120
QY 6469 cgcacccctgcgcgcgcgcctctactctcctcctcctcctcctcctcctcctcctcctc 6528
Db 121 CGCACCCCTGCGCGCGCCCTCTACTCTCTCTGCGGCGCAGCTGCGCCGCTCTGGACCT 180
QY 6529 ccgcgcgcacccagtgacctctacctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6588
Db 181 CGCGCGCCACCCAGTAGTACCTTACTCTCCGCCCGCGGACCATGTGTCGCCCGCGTGG 240
QY 6589 ctccccccacacgcgaagggggcgaaggtctccagagcgaacacagacgtcgtgtttg 6648
Db 241 CTCCCCCACACAGCAAGGGGGCAAGAGGTCTCCAGAGCCAAACAGACGTCGGTCTTGGG 300
QY 6649 tgggtgagcagcgtattgaacctgtgtcccccacccgagggcagcagcagcagcagcag 6708
Db 301 TGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGCGATGACGAGGCCAGGGCA 360
QY 6709 ctcccgagtgctgtgtaccgcgtgtgtaccggtgtgtaccggtgtgtaccggtgtgtacc 6768
Db 361 CTCGCGGAGTGTGTGTACCCGCTGCTGTGTACCGGATGGGGAACAGACAGGCCACAGCAG 420
QY 6769 gatgggtccaaagtctccagggcaacaccagcagcagcagcagcagcagcagcagcagc 6828
Db 421 GATGGGCTCCAAAGTCTCCAGGCAACACCAGCCAGCCGCGCCAGCCTTCTTCAGCAAGCTGAC 480
QY 6829 cgaagcaactccgcctcgtgtcaagtcgaagcaagcaagcaagcaagcaagcaagcaagca 6888
Db 481 CGAGAGCAACTCCGCCATGGTCAAGTCCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 540
QY 6889 ccacacccggaatgagcctgaataataataataataataataataataataataataataata 6948
Db 541 CCACACCCGGAATGAGCCTGAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 600
QY 6949 gcccgccatcaccggaacagg 6969
Db 601 GCCCGCCATCACCAGGACAGG 621

```



```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE     1 (bases 1 to 629)
AUTHORS      NTH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgabbs@email.nih.gov
              Tissue Procurement: DCTD/DTF
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCM803 row: i column: 18
              High quality sequence stop: 629.

FEATURES             Location/Qualifiers
     1..629
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:3945305"
         /clone_lib="NIH_MGC_7"
         /tissue_type="small cell carcinoma"
         /cell_line="MGC3"
         /lab_host="DH10B (phage-resistant)"
         /note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
         EcoRI; CDNA made by oligo-dT priming. Directionally
         cloned into EcoRI/XhoI sites using the following 5'
         adaptor: GGCACGAG(G). Size-selected >500bp for average
         insert size 1.8kb. Library constructed by Ling Hong in
         the laboratory of Gerald M. Rubin (University of
         California, Berkeley) using ZAP-cDNA synthesis kit
         (Stratagene) and Superscript II RT (Life Technologies)."/>
     130 a 181 c 171 g 147 t
BASE COUNT
ORIGIN

```

Query Match	7.2%; Score 613.8; DB 140; Length 629;
Best Local Similarity	99.5%; Pred. No. 5.3e-105;
Matches 626; Conservative	0; Mismatches 2; Indels 1; Gaps 1;
QY 7922	gcctccccgaatgcatttggacaaagttaaactgaagctcgagcccccgcgcctc 7981
Db	
1	GCCTCCCCGNAATGCAATTGGAAACCAAGCTAACTGAGTTCGAGCCCCCGCGCCCTC 60
QY 7982	cctcgcctcccatccccgccttagcgtctbtgacagatgagcgagggcctgtccagcccc 8041
Db	
61	CCTCGCGCTCCCATCCCGGCTTAGCGGCTTGACAGATGGAGCGAGGCGCTGTCCAGGCC 120
QY 8042	cagtcgcctcgttccgctccccacagactccccagccaacagagatgctgagaaccaag 8101
Db	
121	CAGTGCCTCGTTCGGGTCCACAGACTGCCCCAGCCACAGAGATTGCTGGAAACCAAG 180
QY 8102	tcaggccaggttgggcgagacaaaaggccagtgcgccctgggggaaacggatgctccgag 8161
Db	
181	TCAGGCCAGGTGGCGGACAAAAGGGCCAGTGGCGGCTGGGGGGAACGGATGCTCCGAG 240
QY 8162	gactgactgtttttcacacatcgtttccgacgctggtgggaaggaaagcagatgtaa 8221
Db	
241	GACTTGGACTGTTTTTTCACACATCGTTGCCGACGCGGTGGGAAGGAAGACGATGTAA 300
QY 8222	atgatgtgtgtgtttacagggttatattttgatacccttcaatgaatlaattcagatgttt 8281
Db	
301	ATGATGTGTGTGTTTACAGGATATATTTTGTATACCTTCAATGAATTAATTCAGATGTTT 360
QY 8282	tacgaaggaagacttaccagttattactgctcgtgtgtttgatctctgttaccgt 8341
Db	
361	TACGCAAGGAAGACTATTACCAGTATTACTGCTGCTGTCTTTTGATCTTCGCTTACCGT 420
QY 8342	tcagaggcgtgtgcaggccgacgtcggtgaccccatcactcgcaggacacaaagggcg 8401
Db	
421	TCAAGAGCGGTGTGCAGGCGCAAGTCCGTTGACCCCATCACTCCGAGGACCAAGGGGCG 480

Qy	8402	gggagctgtcg-tcaagccgcgcgtgtgtcctcccctccctcccttccttcccttggtggcagaatgaa	8460
Db	481	GGGAGCTGTGGCTCAGCGCCCGCTGTGTCTCCCTCCCTCCCTTCTCTTTGGGCAGAATGAA	540
Qy	8461	ttcgatgcgtattctgtaggccgaattgccaggggttggttatctgtcatcttacaca	8520
Db	541	TTCGATGCGTATTCTGTGGCGGCCCATCTGCGCAGGGTGGTGGTATTCTGTCAATTTACACA	600
Qy	8521	cgtcgttctaataaaaaagcgaattatat	8549
Db	601	CGTCGTTCTAATAAAAAAGCGAATTATAC	629
RESULT	14		
BE264455			
LOCUS	BE264455	650 bp mRNA	EST 13-JUL-2000
DEFINITION	601191770F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3535648 5',		
ACCESSION	BE264455		
VERSION	BE264455.1	GI:9138011	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 650)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcgabps-r@mail.nih.gov Plate: LIFCM216 row: 1 column: 17 High quality sequence stop: 649.		
FEATURES	Location/Qualifiers		
source	1..650		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3535648"		
	/clone_lib="NIH_MGC_7"		
	/tissue_type="small cell carcinoma"		
	/cell_line="MGC3"		
	/lab_host="DHIOB (phage-resistant)"		
	/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	135 a 189 c 197 g 129 t		
ORIGIN			
Query Match	7.1%; Score 609.2; DB 165; Length 650;		
Best Local Similarity	99.1%; Pred. No. 3.9e-104;		
Matches	644; Conservative 0; Mismatches 3; Indels 3; Gaps 3;		
Qy	7723	ttagctgtctaaagcccttaactaagactccgcgccgggtggccctgtgcagaccttac	7782
Db	1	TTGCCTGTCTAAAGCCCTTAACATAGCTCCCGCCCCGGCTGGCCCTGTGCAGACCTTAC	60
Qy	7783	tcagg-ggaattttacctgtgctcggaagggagggaagggccgggggggggcacg	7841
Db	61	TCAGGTGGATGTTTACCTGGTGCTCGGAAGGAGGGAGGGCCGGGGAGGGCAGC	120
Qy	7842	gcaggcgtgtggcacacacacagggcggccaggcggcgccaggaccaccaagcaggatga	7901
Db	121	GCAGGCGGTGTGGCAGCACACGACGAGCGCGCGCAGGGTGGCCA-GGACCCCAAAGCAGGATGA	179
Qy	7902	ccagcacctccacgcgcactgcctccccccaagtcatcttgaaaccaagtctaaactgag	7961



```
Db 180 CCACGACCTCCACGCCACTGCTCCCGAATGCAATTTGGAACCAAAAGTCTAAACTGAG 239
QY 7962 ctgcagccccgcgcctccctccgcctcccatccgcttaagcgtctggacagatga 8021
Db 240 CTCGAGCCCCCGCGCCCTCCCTCCGCCCTCCCATCCCGCTTAGCGCTCTGGACAGATGA 299
QY 8022 cgcagggccctctcagccccccagtcgctcgttccgggtcccccacagactgccccagccaa 8081
Db 300 CGCAGGCCCTCTCCAGCCCCCAGTGGCTCGTTCCGGTCCCCACAGACTGCCCCAGCCAA 359
QY 8082 ctagatttgtgaacaaagtcaggccaggtgggcggagacaaaggccaggtgcggcctg 8141
Db 360 CGAGATTGCTGGAACCAAGTCAGGCCAGGTGGCGGACAAAAGGCCCAAGTGCGGCCCTG 419
QY 8142 gggggaacggtatgctccgagggactgactgtttttttcacacatgcttcccgagcggtg 8201
Db 420 GGGGAACGGATGCTCCGAGGACTGGACTGTTTTTTTTCACACATCGTTGCCCGCGGGTG 479
QY 8202 ggaagaaaggcagatgtaataga-tgtgtgtgtttacaggggtatatattttgataccttc 8260
Db 480 GGAAGAAAGGCAGATGTAATGACTGTGTCGGTTTACAGGGTATATTTTTTGATACCTTC 539
QY 8261 aatgaattaatcagatgttttacgcaagaagagacttaccagttactactcgtctgtg 8320
Db 540 AATGAATTAATTCAGATGTTTTACGCAAGAAAGGACTTACCAGTATTTACTGCTGCTGTG 599
QY 8321 ctttgatctcttaccgttcaagagcggtgtgcagggccgacagtcgg 8370
Db 600 CTTTGTATCTGCTTACCGTTCAAGAGCGGTGTGCAGGCCCGCACAGTCGG 649

RESULT 15
LOCUS BG391632 624 bp mRNA EST 12-MAR-2001
DEFINITION 602417893F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537438 5',
mRNA sequence.
ACCESSION BG391632
VERSION BG391632.1 GI:13285080
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0462 row: 1 column: 23
High quality sequence stop: 623.

FEATURES
Location/Qualifiers
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4537438"
/clone.lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
136 a 222 c 190 g 76 t

BASE COUNT
ORIGIN
```

```
Query Match 7.0%; Score 602; DB 153; Length 624;
Best Local Similarity 99.7%; Pred. No. 8.7e-103;
Matches 624; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 3660 acagcgccatcacataccgcggtccatccaccagggcacgcccagctgacgtctgttaca 3719
Db 1 ACAGGCCCATACATACCGCGGCTCCATCACCCACGCCAGCTGACGTCTCTGTACA 60
QY 3720 aggggaacatcacccaggatcatcgcgagagacagcccgagtcgcttggaccgcgccggg 3779
Db 61 AGGGCACCATCATCACAGGATCATCGCGAGGACAGACGCCGAGTCTGTGGACCGCGCGG 120
QY 3780 aggaagcctgcgccagggccagctcatctacgaaggaagaagggccagctcttctct 3839
Db 121 AGGACAGCTGCTGCCAAGGCCACGCTCATCTACGAAGCAAGAAGGCCACGCTCTTCT 180
QY 3840 atgagggtggcatgtctgtgaccagtgctccaaaggagagacggcagaagcagctcaggac 3899
Db 181 ATGAGGGTGGCATGTCTGTGACCCAGTGTCTCAAGGAGGACGGCAGACAGCTCAGGAC 240
QY 3900 cccccatgagacgcccgcgcccccaagcgcaacctatgacatgagggcgctgggca 3959
Db 241 CCCCCATGAGACGGCGCGCCGCCCAAGCGCACCTATGACATGATGGAGGCGCGGTGGCA 300
QY 3960 gagccatctccagcagcatcgaaggtctcatggcgctgccatcccgccgagcgac 4019
Db 301 GAGCCATCTCTCAGCCAGCATCGAAGGTCTCATGGGCGCTGCCATCCCGCGGAGCGAC 360
QY 4020 acagccccccacacctcaagagcagcaccacatcccgcggtccatcacacaaggatcc 4079
Db 361 ACAGCCCCCACCACCTCAAGAGAGCAGCACACATCCGCGGGTCCATCACACAAGGATCC 420
QY 4080 ctcggtctctacgtgagggcacagggagactacctcgctcgaggagccaagctctaaagc 4139
Db 421 CTCGCTCTTACGTGGAGGCACAGGAGGACTACTCTGCTCGGAGGCAAGCTCTCTAAAGC 480
QY 4140 gggagggaacgctccgccccccacgcccctcaggggacctgaccgagggcctacaagacg 4199
Db 481 GGGAGGGACAGCTCCGCGCCGCCACCGCCCTCAGGGAGCTGACCGAGGCTTACAGACGC 540
QY 4200 agggcctggggccccctgaaagctgaagccggcccatgagggcctgggtggccacggtgaag 4259
Db 541 AGGCCCTGGGCGCCCTGAAGCTGAAGCCGG-CCATGAGGGCCTGGTGGCCACGGTGAAGG 599
QY 4260 agggggcgctccatccatgatc 4285
Db 600 AGGCGGGCGG-TCCATCCATGAGATC 624
```

Search completed: September 8, 2001, 14:32:15  
Job time: 17647 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2001, 09:38:08 ; Search time 10979.9 Seconds  
(without alignments)  
12074.241 Million cell updates/sec

Title: US-09-522-753-4  
Perfect score: 8571  
Sequence: 1 catgtcgggctccacacagc.....caaaaaaaaaaaaaaaaaaa 8571

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_rod:\*  
33: em\_hum1:\*  
34: em\_hum2:\*  
35: em\_hum3:\*  
36: em\_hum4:\*  
37: em\_hum5:\*  
38: em\_hum6:\*  
39: em\_hum7:\*  
40: em\_in:\*  
41: em\_om:\*  
42: em\_or:\*  
43:

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vil2:\*  
60: gb\_vil3:\*  
61: gb\_vil4:\*  
62: gb\_vil5:\*  
63: gb\_vil6:\*  
64: gb\_vil7:\*  
65: gb\_vil8:\*  
66: gb\_vil9:\*  
67: gb\_vil10:\*  
68: gb\_vil11:\*  
69: gb\_vil12:\*  
70: gb\_vil13:\*  
71: gb\_vil14:\*  
72: gb\_vil15:\*  
73: gb\_vil16:\*  
74: gb\_vil17:\*  
75: gb\_vil18:\*  
76: gb\_vil19:\*  
77: gb\_vil20:\*  
78: gb\_vil21:\*  
79: gb\_vil22:\*  
80: gb\_vil23:\*  
81: gb\_vil24:\*  
82: gb\_vil25:\*  
83: gb\_vil26:\*  
84: gb\_vil27:\*  
85: gb\_vil28:\*  
86: gb\_vil29:\*  
87: gb\_vil30:\*  
88: gb\_vil31:\*  
89: gb\_vil32:\*  
90: gb\_vil33:\*  
91: gb\_vil34:\*  
92: gb\_vil35:\*  
93: gb\_vil36:\*  
94: gb\_vil37:\*  
95: gb\_vil38:\*  
96: gb\_vil39:\*  
97: gb\_vil40:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	8541	99.6	8561	89	AF113003
2	8324.6	97.1	8686	89	AF113003 Homo sapi
3	5455	63.6	5989	97	AF125672 Homo sapi
4	5037.2	58.8	8544	94	U37146 Human silen
5	4981.6	58.1	8388	94	AF125671 Mus muscu
6	4378.2	51.1	7465	94	AF113001 Mus muscu
7	2759.8	32.2	2842	91	AF13002 Mus muscu
8	2618.6	30.6	2930	97	BC004326 Homo sapi
					S83390 T3 receptor











Db 2341 AGGCCACCCACCCACACGAGGAGACATCCCGGGGCCCCCATTTGAGCCCCACCCCGGCCTC 2400  
Qy tgaagccacggagcccttaccgccccaccacagcaccctccatcgccctctgacacctctcc 2460  
Db 2401 TGAAGCCACCGGAGCCCTACGCCCCACACAGCACCCCATCGCCCTCTGACCTTCCTCC 2460  
Qy tctgttccccaaaggaggaggaaggagagagacccagcagcagccccccagtgagagagg 2520  
Db 2461 TGTGTTCCCAAGAGGAGAGAGGAGGAGAGACCGACGACGCGCCCTCAGTGAGAGGG 2520  
Qy 2521 ggagagcagaagcccccgctgagagagctgagagctgagagagagagagagagagag 2580  
Db 2521 GGAGGAGCAGAAGCCCCCGCGGCTGAGAGAGTGGACAGTGGACAGAGGAAAGGCGAGGA 2580  
Qy 2581 gccctcaagagcagtgacagaggaagccagagagagagagagagagagagagagagc 2640  
Db 2581 GCCCTCAAGACGAGTGCACGGAGAGAGCCGAGGAGGGCGCGCAAGGCAAGGACGC 2640  
Qy 2641 ggaagccgctgagggccagggcgagggcgctcaaggcagagagagagagagaggggag 2700  
Db 2641 GGAGGCGCTGAGGCCACGGCGGAGGGCGCTCAAGGCAGAGAAAGGAGGGCGGAG 2700  
Qy 2701 cggcagggccaccactgccaagagctcgggagccccccagagagagagagagagagctac 2760  
Db 2701 CGGCAGGGCCACCACTGCCAAGAGTTCGGGGCGCCCCCAGAGACAGCAGTCTCCAGTGTAC 2760  
Qy 2761 ctgcagtcagacagagtgtagagagcggagggcgagagagagagagagagagagagag 2820  
Db 2761 CTGCAAGTCAGACGAGTGTGATGAGCGCGAGGGCGGCGACAGAAACCGGCTGTCTGCC 2820  
Qy 2821 aagggccagctctcaaccggactggcgacccccggggccaatgctcacccccagagagcc 2880  
Db 2821 AAGGCCAGCCTCTCTACCCCGACTGGCGACCCCGGGGCCAATGCTCACCCAGAGGCC 2880  
Qy 2881 actgagactgaagcagctgaagcagcagcagcagcagcagcagcagcagcagcagcagc 2940  
Db 2881 ACTGACCTGAAGCAGCTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2940  
Qy 2941 agtccatgagccccccgggagagcagctcccaagcagcagcagcagcagcagcagcagc 3000  
Db 2941 AGTCCATGAGCCCCCGGAGGAGCGAGCTCCCAAGCAGCAGCTCCCGCCAGCCCCACC 3000  
Qy 3001 gccaccgcaaacctgcagccggagagcgagccccctcagcagcctgcagcagcagcagc 3060  
Db 3001 GCCACCGCAAAACCTGCAGCGGAGAGCGAGCGCCCTCAGCAGCCTGCGCAGCAGCCCGC 3060  
Qy 3061 gggcaagcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3120  
Db 3061 GGGCAAGAGCAGGAGCGCGGACACCCCGCGACAGAGAGGCTTCGAGCGCGAGGCCCA 3120  
Qy 3121 gaagctgctgggagcccccttctgtgacttcgggctgccttccccctgcccccccg 3180  
Db 3121 GAAGCTGCTGGGAGCCCCCTTGTGTGACTTCCGGCTGCGCTTCCCGCTGCCCGCCCG 3180  
Qy 3181 tgagtgatcaagcctcccccgatgccccgaacccctcagccttctctactcagctccacc 3240  
Db 3181 TGAGTGATCAAGGCTCCCGCATGCCCGAGCCGACCCCTCAGCTTCTCTACGCTCCACC 3240  
Qy 3241 tggtaaccactgccccctgggctcccatgacactgccccgccccctctgcgagccacc 3300  
Db 3241 TGGTCAACCACTGGCCCCCTG6GGCTCCCATGACACTGCCCGGCCCTCTGCGCGCCCCACC 3300  
Qy 3301 caccatccccaccgctccccctctctctctgcaagcaccacccagcgtccctcgagag 3360  
Db 3301 CACCATCTCAACCGCTCCCTCATCTCTCTGCAAGCAGCCACCCAGCGTCTCTCGAGAG 3360  
Qy 3361 gcaaatagtgccatctccccaaaggaatgctggtccagctccacgtccccgtactcagagca 3420  
Db 3361 GCAATAGGTGCCATCTCCCAAGGAATTCGGTTCAGCTCCAGCTCCCGTACTCAGAGCA 3420  
Qy 3421 tgcaagcccccgatgggctgctcaccatgggctgccccctgccccatgagccccaaaa 3480  
Db 3421 TGCCAAGCCCCGGTGGGCCCTGTACCATGGGGCTGGCCCTGTGCCCATGGACCCCAAAAA 3480

Qy 3481 gctggcaccttcagcgagtgaaagcagagcagctgtcccccagggggccaggtctgggccc 3540  
Db 3481 GCTGGCACCTTCAGCGAGTGAAGCAGAGCAGCTGTCCCCAGGGGCGCAGGTGGGCC 3540  
Qy 3541 accggagagcctgggggtgcccacagcccagagcgctccgtgtctgagagggagacgtct 3600  
Db 3541 ACCGGAGAGCCTGGGGTGGCCACAGCCACAGGCGCTCCGTGTGAGAGGGACAGCTCT 3600  
Qy 3601 gggctcagttccggggcgaagcattcccaagcacaaggggtgcctcgga 3660  
Db 3601 GGGCTCAGTTCGGGGCGGAAGCATACCAAAGGCATTCCACAGCACACGGGTGCCCTCGGA 3660  
Qy 3661 cagcgccatcataccgcgctccatcaccacagcgacgcaagcagctgacgtccctgtacaa 3720  
Db 3661 CAGCGCATCATACCGCGCTCCATCAACCCACGGCACGCCAGCTGACGTCTCTGTACAA 3720  
Qy 3721 gggcaccatcaccaggatcctcgcgagagacagcccagctgcttggaccgccccggga 3780  
Db 3721 GGGCACCATCACAGGATCATCGCGAGGACAGCCGAGTCTGAGACCGCGCGGGGA 3780  
Qy 3781 ggaagcctgcccgaagggcagctcatctacaaggaagaagggccacgtctcttctcta 3840  
Db 3781 GGACAGCCTGCCCAAGGGCCACGTCATCTACAAAGGCAAGAGGCCACGCTTGTTCCTA 3840  
Qy 3841 tgaggtggtcagctgtgtgacccagctgtccaagagagcagcagagaagcagctcagga 3900  
Db 3841 TGAGGTGGCATGTCTGTGACCCAGTGTCTCCAAAGGAGACGGCAGAGCAGCTCAGGACC 3900  
Qy 3901 ccccatgagagcgccgcccccaagcgcaactatgacatgagagggcgccgtggggcag 3960  
Db 3901 CCCCATGAGAGCGCGCGCCCAAGCGCACCTATGACATGATGAGGGCGCGTGGGCGAG 3960  
Qy 3961 agcatctctcagccagcagctgaaggtctcatgggcccgtgccatccccgcggagcgaca 4020  
Db 3961 AGCCATCTCTCAGCCAGCATCGAAGGTCTCATGGCGGTGCCATCCCGCGCGAGCGACA 4020  
Qy 4021 cagccccaccacactcaagagcagcaccacatccgcggttccatcacacaagagatccc 4080  
Db 4021 CAGCCCCACCCACTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 4080  
Qy 4081 tcggtctctacgtggagcagagggactacctgctgggagggccaaagctctctaaagcg 4140  
Db 4081 TCGGTCTCTACGTGGAGGCACAGGAGGACTACTGCTGCGGGAGCCAAAGCTCTCTAAAGCG 4140  
Qy 4141 ggaagcagcgtccgcccccaacgcgctcagggacctgaccagagggcctacaagagcga 4200  
Db 4141 GGAGGGCACGCTCTCGCCCCCACCCTCACGGGACCTTGACCAGGCGCTTCAAGAGCGCA 4200  
Qy 4201 ggcctgggccccctgaagctgaagcggcccatgagggcctggtggccaggtgaagga 4260  
Db 4201 GGCCCTGGGCCCCCTGAAGCTGAAGCGGCCCATGAGGGCTGTGTGCCACAGGTGAAGGA 4260  
Qy 4261 ggcggcgctccatccatgagatccccgcgcgagagctgcggcacagccccagctgccc 4320  
Db 4261 GCGCGCGCTTCCATCCATGATGATCCCGCGGAGAGTGTGGGCACACAGCCCGAGCTGCC 4320  
Qy 4321 cctggcccccgcgctcgaagagggctccatcagcaggggaacccccgtccaagtaaga 4380  
Db 4321 CCTGGCCCCGCGCGCTCAAGAGAGGCTCCATCACCGAGGGCACCCCGCTCAAGTAGCA 4380  
Qy 4381 caecggcgcttccaccactggtccaaaaaacgacgctacgtccctcatcggcagcccc 4440  
Db 4381 CACCGCGCGTCCACCACTGGCTCCAAAAGCAGCAGTACGCTCCCTCATCGCAGCCCC 4440  
Qy 4441 cggcgagcgttccccaccctgacccccctggaatgtatggccgagcccccggtcactgga 4500  
Db 4441 CGGCGGAGCTTCCACCCGTCACCGCTGATGTGATGGCCGACGCGCGGCGACTGGA 4500  
Qy 4501 acgtgctgctacagagagagcctgaagaccgagcagcagcagcagcagcagcagcagc 4560  
Db 4501 ACGTGCCTGTCTACGAGGAGAGCTTGAAGAGCGCGGCCAGGGACCCAGCTCGGGGG 4560







Db 6721 TGCTGACCCGCTGCTGTACCGGGATGGGAAACAGACAGCGGCCAGCAGAGATGGGCTCAA 6780  
Qy 6781 gtctcaggcaacaccagccagccagctcttccagcaagctgacgagagcaactc 6840  
Db 6781 GTCTCAGGCACACAGCCAGCCGCCAGCCCTTCTTCAGCAAGCTGACCGAGAGCACTC 6840  
Qy 6841 cgccatgttcaagttccaagaagcaagagatcaacaagaagctgaaaccccccaaccggaa 6900  
Db 6841 CGCCATGTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCTGAACACCCACACCGGAA 6900  
Qy 6901 ttagcctgaatacaataatcagcagcctggagcagagatcttcaatatgcccgccatcac 6960  
Db 6901 TGAGCCTGAATACAAATATCAGCAGCCTGGGACGGAGATCTTCAATATGCCCGCATCAC 6960  
Qy 6961 cgaacagccttatgacctatagaagcagcggtgaggaacatgcagcaccacaat 7020  
Db 6961 CGGAACAGCCTTATGACCTATAGAAGCAGCGCGGTGCAGGAACATGCGACCAACAT 7020  
Qy 7021 ggggctggaggccataattagaaggcactcatgggttaatatgaccagtggaagagtc 7080  
Db 7021 GGGGCTGGAGGCCATAATTAGAAGGCACCTCATGGGTAATATGACCAAGTGGGAAGATC 7080  
Qy 7081 cccgcgctcagcgccaatgcttttaaccctctgtaacccctgagtcagcctgcccgtgc 7140  
Db 7081 CCCGCGCTCAGCGCAATGCTTTTAACCTCTGAATGCCAGTGCAGCCTGCCCGCTGC 7140  
Qy 7141 tatgccataacgctgctgacgagcagtgagtcacacacactcacctgcagcagctggcg 7200  
Db 7141 TATGCCATATACCGTGTGACGACGAGTGACACACACTCACCTGCCAGGTGGCGG 7200  
Qy 7201 cgggaagccaggctctctgagaccagcagcgagcgaagccaaagtcgcccgcccg 7260  
Db 7201 CGGAAGGCCAAGGCTCTGGCAGACCCAGCAGCGCGAAGGCCAAGTCCCGCGCCCGG 7260  
Qy 7261 cctggcatctgggaccggccaacctctgtctcctcagtgcaactggagggagagctgaa 7320  
Db 7261 CTTGGCATCTGGGGACCGCCGACCCCTCTGTCTCTCTCACTGCGAGGAGAGCTGAA 7320  
Qy 7321 ccgcggagcgcctcaccacccgctggtggagagcagccctgctgcgaggttcac 7380  
Db 7321 CGCGCGAGCGGCTCACCAACCGGCTGGGAGACAGGCCCTCTGTCGGCAGGTTCCAC 7380  
Qy 7381 gccattccccataacccccctgataatcggtcagcgcggtgtcatgggttcccacc 7440  
Db 7381 GCCATTCCCTTACAAACCCCTGATCATCGGCTGCAGCGGGTGTATGGCTTCCACC 7440  
Qy 7441 cccacgggctcccgcgggcagcgccctgctggcccccaacacgcgtggagca 7500  
Db 7441 CCCACGGGCTCCCGCGGGCAGCGGGCCCTCGCTGGCCCCCACACGCTGGGACGA 7500  
Qy 7501 ggagcccaagcactgctctgcagtagacagacactctccagacgagtgactcag 7560  
Db 7501 GGAGCCCAAGCCACTGCTCTGCTCGCAGTAGAGACACTCTCCACACCGAGTACTCAG 7560  
Qy 7561 aacagggcgggggggcgggcggtgcuersperlugtcagtcgccagcagccacaggaa 7620  
Db 7561 AACAGGCGGGGGGGGGGGCGG-----TGTCAGGTCCCGAGGACACAGAA 7610  
Qy 7621 cggcctcagagagcgggcggtgcgagactcccccaacaaaggaagagccctgagtc 7680  
Db 7611 CGGCCCTCAGAGGGGGGGGGCTGCCGACTCCGCCAACCAAGAAAGAGGCCCTGAGTC 7670  
Qy 7681 cgcctggcctccatccatctctcgtccagagcgcgcctctgctgctgtctaaagcctt 7740  
Db 7671 CGCCTGCGCCTCCATCCATCTGCTCGTCCAGACGGGATCTTGGCTGTCTAAGCCTT 7730  
Qy 7741 aactaagactccccgggctggcctgtgacagacttactcaggggatgtttacctg 7800  
Db 7731 AACTAAGACTCCCGCCCGGGCTGGCCCTGTGCGACACTTACTCAGGGGATGTTTACCTG 7790  
Qy 7801 gtgctcggagagagagggcgggcgagggggcgagggcgagggcggtgtggcagcac 7860  
Db 7791 GTGCTCGGGAAGGGGGAAGGGGGCGGGGAGGGGGCACGGCGGCGGTGTGGCAGCCAC 7850

Qy 7861 acacggcgccagggcgccagggaccacaaagcaggatgaccagcaccctccacggcac 7920  
Db 7861 ACACAGGGCGCCAGGGCGCCAGGGACCCAAAGCAGGATGACACGACCTCCACGCCAC 7910  
Qy 7921 tgctccccccaatgctatttggaaacaaagtctaaactgagctcgagcccccgccct 7980  
Db 7911 TGCTCCCCCGAATGCAATTTGGAACCAAGTCTAACTGAGCTCGCAGCCCCCGGCCCT 7970  
Qy 7981 cctcgcgctcccatcccgcttagcgctctggacagatggagcagggccctgtccagccc 8040  
Db 7971 CCTCCGCTCCCATCCGCTTAGCGCTCTGACAGATGGACGAGGCCCTGTCCAGGCC 8030  
Qy 8041 ccagtgcgctgcttccggtccccacagactccccacagactccccacagagattgtggaaacaa 8100  
Db 8031 CCAGTGCCTCTGTTCCGCTCCCGACAGACTGCCCGACGCAACAGAGATGCTGGGAACCAA 8090  
Qy 8101 gtcagggcaggtggcgacaaaaggcggtgcggtgcggtggggaaggaaggaagatgta 8220  
Db 8091 CTCAGGCCAGCTGGCGGACAAAAGGGCCAGTGGGCTGGGGGACGATGCTCCGA 8150  
Qy 8161 ggaactggactgttttttcacacatgcttgcgcagcggtggggaaggaaggaagatgta 8220  
Db 8151 GGACTGGACTGTTTTTTTCACACATCGTTGCCGACGCGTGGGAAGGAAGGACATGTA 8210  
Qy 8221 aatgatgtgtgtttacagggatatttttgataccttcaatgaatattcagatgtt 8280  
Db 8211 AATGATGTGTGTTCACAGGTATATTTTGTATACCTTCAATGAATTAATTCAGATGT 8270  
Qy 8281 ttacgcaagggaagacttaccagctattactgctgtgctgttttgcctgtcttaccg 8340  
Db 8271 TTACGCAAGGAAGGACTTACCCAGTATTACTGCTGCTGTGCTTTTGTATCTCTGCTTACC 8330  
Qy 8341 ttcaagaggcggtgcagggcgacagtcggtgaccccaactcactgcagggaccagggggc 8400  
Db 8331 TTCAAGAGCGGTGTCAGGGCGCAGTCGGTGACCCCATCATCGCAGGACCAAGGGGCG 8390  
Qy 8401 ggggactgctcgtcacgccccgctggtccctccctccctccctccctggcgagaatgaa 8460  
Db 8391 GGGGACTGCTGCTCAGCGCCCGCTGTGTCCTCCCTCCCTCCCTTCCCTGGGCAAGTAA 8450  
Qy 8461 ttcatgctgattctgtggcgccatttgcaggggtggtgattctgtcattacaca 8520  
Db 8451 TTCGATGCTATTCTGTGGCGCCATTTGCGAGGGTGGTGTATCTGTCTATTACACA 8510  
Qy 8521 cgtgcttctaataaaagcgaattataactccaaaaaatactccaaaaaatactccaaaaa 8571  
Db 8511 CGTCGTTCTAATTAAGAGCGAATTATATCTCCAAAAAATACTCCAAAAAATACTCCAAAAA 8561

## RESULT 2

AF125672 8686 bp mRNA PRI 04-APR-1999  
LOCUS  
DEFINITION Homo sapiens silencing mediator of retinoic acid and thyroid hormone receptor extended isoform (SMRTE) mRNA, complete cds.  
ACCESSION AF125672  
VERSION AF125672.1 GI:4559297  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 8686)  
AUTHORS Park, E.J., Schroen, D.J., Yang, M., Li, H., Li, L., and Chen, J.D.  
TITLE SMRTE, a silencing mediator for retinoid and thyroid hormone receptors-extended isoform that is more related to the nuclear receptor corepressor  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)  
MEDLINE 99199215  
REFERENCE 2 (bases 1 to 8686)  
AUTHORS Chen, J.D.  
TITLE Direct Submission  
JOURNAL Submitted (03-FEB-1999); Pharmacology and Molecular Toxicology,











Qy 3337 caagcaccacagctctcagagagcaaaatagtgccatctcccaagaaatgtcgttcca 3396  
Dy 3462 CAAGCACCACCGCTCTCGAGAGGCAAAATAGTGCCATCTCCCAAGGAATGTCCGGTCCA 3521  
Qy 3397 gctccacgtcccgtaactcaagagcatgccaaagccccgggtggccctgtcaccatggggct 3456  
Dy 3522 GCTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCGGTGGGGCTGTACCATGGGGCT 3581  
Qy 3457 gccctgcccattgaccccaaaagctggccacctcagcggagtgaagcaggagcagct 3516  
Dy 3582 GCCCTGCCCATGACCCCAAAAGCTGGCACCTTCAGCGGAGTGAAGAGAGGAGCCT 3641  
Qy 3517 gtccccacggggccaggctggccacccggagagcctgggggtgcccacagccccaggagc 3576  
Dy 3642 GTCCCCACGGGGCCAGCTGGGCCACCGAGAGGCTGGGGGTGCCACAGCCCAGGAGGC 3701  
Qy 3577 gtcctgctgagaagggacagctctggctcagttccgggtccggggaagcatcaccaaagcatt 3636  
Dy 3702 GTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGGCGGAAGCATCACAAAGGCAT 3761  
Qy 3637 tccagcacacgggtgcctcggagcagcgcacatcacatccgcgctccatcacccacgg 3696  
Dy 3762 TCCAGGACACGGGTGCCCTCGGACAGCGCCATCACATACCGCGGCTCCATCACCCACGG 3821  
Qy 3697 cagccagctgacgtctctgtacaagggcaccatcacaggatcatcgggagagcagccc 3756  
Dy 3822 CACGCCAGCTGACCTCTGTACAAAGGCAACCATCACAGGATCATCGCGAGAGCAGCCC 3881  
Qy 3757 gagtctgtgaccgcggccggaggagcctgccaaagggccacgtctctcagaagg 3816  
Dy 3882 GAGTCTGCTTGACCGCGCGGGAGGACAGCTGCCCCAAGGGCCACGTCATCTACGAAGG 3941  
Qy 3817 caaagagggccactgtctctatgaggtggcattctgtgaccagtgctccaagga 3876  
Dy 3942 CAAGAAGGGCCACTGTCTCTATGAGGGTGGCATGCTGTGACCCAGTCTCCAAGGA 4001  
Qy 3877 ggaaggcagaagcagctcaggacccccccatgagacggcgcccccgaagcgacactatga 3936  
Dy 4002 GGACGGCAGAAGCAGCTCAGGAACCCCCCATGAGACGGCGCCCCCAAGCGCACCTATGA 4061  
Qy 3937 catgatggaggggccgctgagagagccatctctcagcagcatcgaaggctcctatggg 3996  
Dy 4062 CATGATGGAGGGCCGCTGGGCAAGCCATCTCTCAAGCAGCATCGAAGGTCTCATGGG 4121  
Qy 3997 ccgtgcatcccgccgagcgacacagccccccacactcaaaagcagcacacacatccg 4056  
Dy 4122 CCGTGCCATCCCGCGGAGGACACAGCCCCCACCCTCAAGAGCAGCACCATCCG 4181  
Qy 4057 cgggtccatcacacaaggatccctcgtctctacgtgagggacagaggactacctcg 4116  
Dy 4182 CGGCTCCATCACACAAGGATCCCTCGTCTACGTGGAGGCACAGAGGACTACCTGG 4241  
Qy 4117 tcggaggccaagctctaaaggggggggaagcctccgccccccacccgcccacaggga 4176  
Dy 4242 TCGGGAGGCCAAGCTCTAAAGCGGGAGGGCAGCCTCCGCCCCCAGCGCCCTCACGGGA 4301  
Qy 4177 cctgaccagagcctacaagcagcagccctggcccccctgaagctgaagcggcccatga 4236  
Dy 4302 CCTGACCAGGCTCTAACAGCGCAGGCCCCCTGGGCCCTGAAGCTGAAGCGGGCCCATGA 4361  
Qy 4237 gggcctggtgcccagctgaaggaggcgggcccgtctccatccatgagatcccgcgcgagga 4296  
Dy 4362 GGGCTGTGGCCACAGCTGAAGAGGCGGGCCGCTCCATCCATCAGATCCCGCGCGAGGA 4421  
Qy 4297 gctgcggcaacgcccagctgccccggcccgccgctcgaaggagggtccatccatc 4356  
Dy 4422 GCTCGGCGACACGCCGAGCTGCCCTGGCCCGCGCGGCTCAAGGAGGGCTCCATCAC 4481  
Qy 4357 qcaggcaccctcgaaglacacacggcgctccaccactgctccaaaagacga 4416  
Dy 4482 GCAGGACCCCGCTCAAGTACGACACCGGGCGCTCCACCCTGGCTCCCAAAAGGACGA 4541

Qy 4417 cgtacgtccctctatcgcgagccccgcggcagcttccccacccgtgacccccgctggatgt 4476  
Dy 4542 CGTACGCTCCCTCATCGCAGCCCGCGCGGAGCTTCCACCCGTGCACCCCGTGGATGT 4601  
Qy 4477 gatggcgaagccccgcgcaactggaaagctgctgctacagagagagcctgaagagccggcc 4536  
Dy 4602 GATGGCGAGACCCCGGGCAGCTGGAAACGTCCTGTACAGGAGAGCCTGAAGAGCGGGCC 4661  
Qy 4537 agggaccgcgaagcagctcggggggtccattgcgcgcgcccccgggtcattgttcctga 4596  
Dy 4662 AGGACCGCCAGCAGCTCGGGGGGCTCCATTGGCGCGCGCCCGGTCATTGTGCTCTGA 4721  
Qy 4597 gctgggtaagccgcggagagcccccttgacctatgagaccacgggggcaacctttggccgg 4656  
Dy 4722 GCTGGGTAAAGCCGCGGAGAGGCCCTTGACCTATGAGGACACAGGGGCACTTGTGCGG 4781  
Qy 4657 ccacctcccaagaggttcgcccgtgacctgagggagccccgcgcgcctgcaggaagg 4716  
Dy 4782 CCACCTCCACAGAGGTTTCGCCCTGACCATCGGGAGCCCCACGCGCGCTGCAGGAGGG 4841  
Qy 4717 cagctttcgtccagcaagggcatcccagggaccgaaagctgagctcgagcctcgtgagat 4776  
Dy 4842 CAGCTTTCTGTCACGAAGGATCCAGGACCGAAAGCTGAGCTGACGCGCTCTGAGAT 4901  
Qy 4777 cgcaagtcctccgcagcagcagctgcccggagcaccacacacccccatctcgccctatga 4836  
Dy 4902 CGCAAGTTCGCCGACAGCAGCCGTGCCGAGCACCACCCACACCCCATCTCGCCCTATGA 4961  
Qy 4837 gcaactgcttgggggtgagtgcggtggaactgtatcgagcgaacatccccctggcctt 4896  
Dy 4962 GCACCTGCTTGGGGCGGTGAGTGGCTGGACCTGTATGCGAGCCACATCCCGCTGGCCTT 5021  
Qy 4897 cgacccacactccataccccgcgcatccctctgagcagcagccctgctactactacc 4956  
Dy 5022 CGACCCACCTTCATACCCCCCGGCATCCCTCTGACGACGCGCTGCTACTACTCTGCC 5081  
Qy 4957 ccgacactggcccccaacccactaccgcacactgaccacacccactcactcactcgcgg 5016  
Dy 5082 CCGACACCTGGCCCCCAACCCACCTACCGCAGCTGCTACCCACCCCTACCTCATCCGCGG 5141  
Qy 5017 ctaccccgacaagcgcgctggagaacccggcagaccatcatcaatgactacatcacctc 5076  
Dy 5142 CTACCCCGACAGCGCGCGCTGGAGAACCGGACAGACCATCATCAATGACTACATCACCTC 5201  
Qy 5077 qcagcagatgcaccacaacacgcgcaccccatgcccagcgagctgatatgctgagggg 5136  
Dy 5202 GCAGCATGTGACCAACACAGCGCCACCGCATGGCCCGCAGCAGCTGATATGCTGAGGGG 5261  
Qy 5137 cctctcgcccgagctcctcgtggcactcaactacgctcggggtcccccaggagcactcat 5196  
Dy 5262 CCTCTGCGCCCGGAGTCTCTCGCTGGCATCAACTACGCTGCGGGTCCCGGAGGCTCAT 5321  
Qy 5197 cgaactgtcccaagtgcacaacctgctgtgctggtgcccgcagacacagcagccccagc 5256  
Dy 5322 CGACCTGTCTCCAAAGTGCACACCTGCTGTGCTGTGTCGCCCGCGACACCGAGCACCCACG 5381  
Qy 5257 caccgcatgaccgccttgcctacctccccacgcgcgcagcccttcagcagcgcca 5316  
Dy 5382 CACGCCCATGACCGCTTGTCTACTCTCCACCGCGCCCGCCTTCAGACGCGCGCA 5441  
Qy 5317 cagcagctccccactctccccagagggtcccaacacttgacaaaacacacacacagctc 5376  
Dy 5442 CAGCAGTCCCCACTCTCCCGCAGGAGGTCCAACACTTGACAAAAACCAACACACACGTC 5501  
Qy 5377 ctgctcgagggagagagaccggggtcgagagcgggaccggggtcggggagcgggaaaaa 5436  
Dy 5502 CTCGTCCGAGCGGAGCGAGACCGGGATCGAGAGCGGGATCGGGAGCGGGGAAAA 5561  
Qy 5437 gtcctactcactcagtcaccacagcggtggagcagcagccatctcggagacctgtacaga 5496  
Dy 5562 GTCCATCTCTACGTCCACCACGAGCGGTGGAGACGACCCCATCTGGAGACCTGGTACAGA 5621  
Qy 5497 gcagagcagcggcagcagcgagcggcggggtgggggggagcagcagcgccccgc 5556



Db 5622 GCAGAGAGCGGACAGCGGAGCGGGGGGGTGGGGGACAGCAGCGCGCCCGC 5681  
Qy 5557 ctccactcccatgccacacagcaactcgcccatctccctcggaccacagatgacctcca 5616  
Db 5682 CTCACCTCCATGCCATGCCACACAGCACTCCGCCATCTCCCTCGGACCCAGATGCCCTCCA 5741  
Qy 5617 gcagagacccagtggtgtttcacacacacagggatgaaggggtatcatcacgcgtgtggagcc 5676  
Db 5742 GCAGAGACCCAGTGTGCTTCACACACAGGATGAGGGGTATCATCACCGTGTGGAGCC 5801  
Qy 5677 cagcaagcccagtgctctaggtccactccactccactccctccctccctccctccctcc 5736  
Db 5802 CAGCAAGCCACCGTCTGAGGTCCACCTCCACCTCCCTCCACCTCCCTCCACCTCCCT 5861  
Qy 5737 attccacactgccacacactgccacactggcggaacctcgatgggtctacctaccct 5796  
Db 5862 ATTCCCACTGCCACCTGCCACTGGGGGGACCTCTGATGGGGTCTACCTTACCT 5921  
Qy 5797 catgagccctcttctgtcccaagagggccccgggtcgccggccagagcgcccg 5856  
Db 5922 CATGGAGCCGCTCTGTGCTGCCAAGAGGCCGCCGGGTGCCGGCCAGAGGCCGCCG 5981  
Qy 5857 agcagacaccgcatgcttctctctcccaagccccccagccccgctccgggtggagccgc 5916  
Db 5982 AGCAGACACCGGCCATGCTCTCTGCCAAGCCGCCAGCCCGCTCCGGGCTGGAGCCCGC 6041  
Qy 5917 ctctccccagcaaggtcgagccccggccctctctctctctctctctctctctctct 5976  
Db 6042 CTCCTCCCCAGCAAGGCTCGGAGCCCCGGCCCCCTAGTGTCTCTCTCTCTCTCTCT 6101  
Qy 5977 caccatcgccgcacccctcgaaagaaactcgcaactcaccacgcagcccgagcccgcc 6036  
Db 6102 CACCATCGCCGCACTCTGAGAGAACTTGCACCTCAACACGCCAGCCCGGAGCCGCC 6161  
Qy 6037 ggcgcaactcctcgctcgaccccgaccccggaagaaactcaaaacttaaccctttc 6096  
Db 6162 GCGCGCACTGCTCGGCTCGGACCCGACCGCGGGAAGAACTCAAAAGTAAACCTTTTC 6221  
Qy 6097 catcaggaactgaactccgtctctgtgtttacacggcagcagctacagcccggaag 6156  
Db 6222 CATCCAGAACTGGAACTCCGTTCTGTGGGTACCCAGGCACTACAGCCCGGAAG 6281  
Qy 6157 ggtgagccctcgacccctgtgactcaccagctctgaccacagcaaggggtccca 6216  
Db 6282 GGTGAGCCCTCAGCCCTGTGAGTCAACCCAGTCTGACCCACGACAGGGGCTCCCA 6341  
Qy 6217 gcaactggaagactcgacaagaccactgaggggagctgpgggcccaagcgacag 6276  
Db 6342 GCACCTGGAAGAGTTCGAAAGAGCCACCTGGAGGGGAGCTGGCGCCCAAGAGCCAGG 6401  
Qy 6277 cccgctgaagcttgaggggagggccgacactccacacactcgccgctgacctgagag 6336  
Db 6402 CCCGTGAAGCTTGGGGGAGGGCCGCCACTCCACACCTGGGGGCTGCTTGAGAG 6461  
Qy 6337 ccagccctcgccagccgctgtccagacggcccgaggggtcaaaagtccaccagcggt 6396  
Db 6462 CCAGCCCTCGTCCAGCCGCTGCTCCAGACCGCCAGGGGTCAAAAGTCAACGAGGGT 6521  
Qy 6397 ggtcaacctgcccagacatacagtgaggtcaacacagagactacacccggacacccc 6456  
Db 6522 GGTCACTGTCGCCAGCAATCAGTGAAGTCACTGAGGTATCACAGAGACTACACCCGCCACCC 6581  
Qy 6457 acagagctagcagccctgcccgcctccctctactctctctctctctctctctctctct 6516  
Db 6582 ACAGAGCTCAGCGACCCCTGCGCGCCGCCCTCTACTCTCTTCCCTGGGGCCAGTGGCC 6641  
Qy 6517 cgtctggaactccgcccacacacagtgactctctactctccgccccggaccatggtgc 6576  
Db 6642 GGTCTGGACCTCGCGGCCACCCAGTACCTTACTCTCCGCCGCCGGACCATGTGTC 6701  
Qy 6577 cccggccgctggtccccccacagcgagggggaaggtctccagagcacaacagag 6636  
|||||

Db 6702 CCCGCCCGTGGCTCCCGCCACAGCGAAGGGGGAAGAGTCTCCAGAGCCAAACAAGAC 6761  
Qy 6637 gtcggtttggtggtggtgaggaaggtattgaacctgtgtccccacccggagggcatgac 6696  
Db 6762 GTCGCTCTGGGTGTTGGTGGAGAGCGGTATTGAACCTGTGTCCCGACCGGAGGCGATGAC 6821  
Qy 6697 ggagccagggcactccccggagtgctgtgtaccgctgtgtaccggggtggtgggaacagac 6756  
Db 6822 GGAGCCAGGGCACTCCCGGAGTGTGTGTACCGCTGTGTACCGGGATGGGGACACAG 6881  
Qy 6757 ggagccagcaggatgggttccaggttccaggcaacaccagccagccgagcctctt 6816  
Db 6882 GGAGCCAGCAGGATGGCTCCAAAGTCTCCAGGCAACACACAGCCAGCCGCGCCTCTT 6941  
Qy 6817 cagaagctgaccagagcaactccgccaatggtcaagtccaagaagcaagatcaacaa 6876  
Db 6942 CAGCAAGCTGACCGAGAGAACTCCGCCATGGTCAAGTCCAAGAAAGCAAGAGATCAACAA 7001  
Qy 6877 gaagctgaacacccac 6936  
Db 7002 GAAGCTGAACACCCAC 7061  
Qy 6937 gatctcaatatgcccgcatacccggaacaggccttatgacctatagaagccagggcgt 6996  
Db 7062 GATCTTCAATATGCCCGCATACCGGAACAGGCTTATGACCTATAGAAGCCAGGCGT 7121  
Qy 6997 qcaggaactcggcagcaccacacacacacacacacacacacacacacacacacacac 7056  
Db 7122 GCAGGACATGCCAGACCAACATGGGGCTGGAGGCCATTAATTTAGAAAGGACACTCATGG 7181  
Qy 7057 taatatgaccagtggaagagtgcccgccgtcagcgcccaatgcttttaacccctctgaa 7116  
Db 7182 TAATATGACCACTGGGAAGAGTCCCGCCGCTCAGCGCCAACTGCTTTTAAACCTCTGAA 7241  
Qy 7117 tgcagtgccagcctgcccgtgtatgtccataacccgtgtgtgacgagcagagtgacca 7176  
Db 7242 TSCAGTGTCCAGCTGCGCGCTGTGTATGCCATTAACCCCTGCTGACGAGCGAGTGACCA 7301  
Qy 7177 cacactcactcgcaggtgaggggaggaagcacaagctctctgacagaccagcagcgg 7236  
Db 7302 CACACTCACTTCGCGAGTGGCGGGAGGCAAGGCTCTGTGACAGCCAGCAGCGCG 7361  
Qy 7237 aaaagccaagtccccggccccggcctggcactctgggagccggccacccctctctctc 7296  
Db 7362 AAAAGCCAAGTCCCGCGCCCGCGCTGGCATCTGGGACCGCCACCTCTGTCTCTCTC 7421  
Qy 7297 agtgcactcggagggagactgcaacccgagcagcgcgtcaccacacccgctgtggagga 7356  
Db 7422 AGTGCACTCGGAGGAGACTGCAACCGCGGACGCGCTCAACCAACCGCTGTGGGAGGA 7481  
Qy 7357 cagccctcgtcccgaggttccacgccaattccctcaacccctgatcatgagcgtgca 7416  
Db 7482 CAGCCCTCTGTCGCGAGGTTCCAGCCATTCCTCCCTACAAACCCCTGATCATGCGGCTCA 7541  
Qy 7417 ggcgggtgctatggtctccccacccccacccggcctcccgcggggagcagggccctcgc 7476  
Db 7542 GCGGGGTCTATGCTTCCCGCCACCCAGCGGCTCCCGCGGCGAGCGGCGCCCTCGC 7601  
Qy 7477 tggccccac 7536  
Db 7602 TGGCCCCAC 7661  
Qy 7537 actctcagacagcagtgactcagaacagggcggggggggggggggggggggggggggg 7596  
Db 7662 ACTCTCGACAGCGAGTACTCAAGACAGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGG 7711  
Qy 7597 tcaggtcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 7656  
Db 7712 TCAGTCCCGAGCGACACAGGACCGCCCTGCAGGAGCGGGCGGCTGCCGACTCCCGCC 7771  
Qy 7657 aaccaaggaaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 7716  
Db 7772 AACCAAGGAAGGAGCCCTGAGTCCGCTGCGCTTCCATCCATCTGCTCCGCTCCAGAGCGG 7831



Qy	7717	gcattctgctgtcctaagcccttaactaaagactccccgcccccggcgccctgtgcaga	7776
Db	7832	GCATCTTGGCTGCTCTAAAGCCTTAACTAAGACTCCCGCCCCGGGCTGGCCCTGTGCAGA	7891
Qy	7777	ccttactcagggaatgtttacctgtgtctcgggaaggaggaaggccggggaagggg	7836
Db	7892	CCTTACTCAGGGGATGTTTACCTGTGTGTCGGGAAGGAGGGGAAGGGGCCGGGAGGGG	7951
Qy	7837	gcacgcgcagcgttgtgcagccacacacagcgcgcgcagggcgccagggcgaccacaagcacg	7896
Db	7952	GCACGCGAGCGTGTCGACGCCACACACAGGCGGCGCACGGGCGGCCAGGCCAACAGCAG	8011
Qy	7897	gatgaccaagcaacctccaagccactgcctccccgaatgcatattggaaaccaaagtctaaa	7956
Db	8012	GATGACCACGCACCTCCAGCCACTGCCTCCC GCCGAATGCTATTGGAAACCAAGTCTAAA	8071
Qy	7957	ctgaegtgcagccccgcgcgcctccctccgcctcccatacccgcttagcgtcttgacaag	8016
Db	8072	CTGAGCTCGCAGCCCCCGCGCTCCCTCCGCGCTCCCATCCGCTTAGCGCTCTGSGACAG	8131
Qy	8017	atggacgcagcctctccagccccagtcgcctcgttcccggtcccccacacagactgcacca	8076
Db	8132	ATGGACGCGAGCCCTGTCCAGCCCCCAGTCGCTCGTTCGGTCCCCACACAGACTGCCCCA	8191
Qy	8077	gccaacgagattgtgtgaaaccaaactcagggccaggtggcgcgacaaaggccaggtgcg	8136
Db	8192	GCCAAACGAGATTGCTGAAACCAAAGTCAGGCCAGGTGGCGGCACAAAGGSCCAGGTGCG	8251
Qy	8137	gcctgggggggaacgagatgtcgcgagactggactgttttttttacacatatogttgcgcag	8196
Db	8252	GCCTGGGGGGAACGGATGCTCCGAGGACTGGACTGTTTTTTTTCACATCGTTGCCGCAG	8311
Qy	8197	cgttggaaggaaagcagatgtaaagatgtgttggtttcacaggggtatatattttgatac	8256
Db	8312	CGTGGGAAGGAAGCCAGATGTAATGATGTGTGTGTTTACAGGGGTATATTTTGTATAC	8371
Qy	8257	cttcaatgaattaattcagatgttttacgaagggaaggaacttaccagttactactgtgc	8316
Db	8372	CTTCAATGAATTAATTCAGATGTTTTACGAAGGAAGACTTACCAGTATTACTGCTGC	8431
Qy	8317	tgtgcttttgatctctgtctacggttcaagaggcgtgtgcagggccgacagtcggtgaccc	8376
Db	8432	TGTGCTTTTGTATCTCTGCTTACCGTTCAGAGGCGTGTGCAAGCGCCAGCATGCGGTGACCC	8491
Qy	8377	catcactccagacacaaaggcgaggactgctcgtcacgcccgctgtgtctccctc	8436
Db	8492	CATCCTCGGAGACCAAGGGCGGGGACTGCTGCTCACGCCCGCTGTGCTCTCCCTC	8551
Qy	8437	cctccctctcttggcagaatcgaatcgaatcgtattctgtggcgccatttgcgcaggg	8496
Db	8552	CTCCCTTCTTGGGCAGAAATGAATTCGATCGTATTCTGTGGCGGCCAATTTGGCGAGG	8611
Qy	8497	tgtgttatctgtcatattcacacgctggttctaattaaaaagcgaattatactccaaaa	8556
Db	8612	TGTTGTATTCTGTCTATTTACACAGTCGTTCTTAATTTAAAAAGGAATTATATCTCCAAA	8671
Qy	8557	aaaaaaaaaaaaaa 8571	
Db	8672	AAAAAAAAAAAAAAAA 8686	
<hr/>			
RESULT	3		
HSU37146			
LOCUS	HSU37146	5989 bp	mRNA PRI 31-OCT-1995
DEFINITION	Human silencing mediator of retinoid and thyroid hormone action (SMRT) mRNA, complete cds.		
ACCESSION	U37146		
VERSION	U37146		
KEYWORDS	U37146.1 GI:1045654		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		











QY 7579 gggcggtteuatersperlugtcaagtcctccagcgagccacaggaagccctgcagggagcg 7638  
||||| : |||||  
Db 5007 GGGCGG-----TGTCAGGTCCAGCAGGACACAGGACGCCCTGCAGGAGCGG 5056  
  
QY 7639 ggggtgcgactcccccaaccagaagagcccttgatcgctgcgctcccatcca 7698  
||||| : |||||  
Db 5057 GGGGTGCGGACTCCCAACCAAGGAAGAGCCCTGAGTCGCGCTCGCTCCATCCA 5116  
  
QY 7699 tctgtccgtcagagcccgcatctgctgtcttaagccttaactaagactccgcgcc 7758  
||||| : |||||  
Db 5117 TGTGTGCGTCCAGAGCGCGCATCTTGCTGCTCTAAAGCCTTAACAGACTCCCGCCCC 5176  
  
QY 7759 gggctggccctgtcagacctactcaggggatgtttacctgtgtcgtcggaagaggg 7818  
||||| : |||||  
Db 5177 GGGCTGGCCCTGTGCAGACCTTACTCAGGGGATGTTTACCTGGTCTCGGAAGGAGGG 5236  
  
QY 7819 gaagggcggggaggggagcgagcggtgtgtgagccacacacagcgccgagggcg 7878  
||||| : |||||  
Db 5237 GAAGGGCGGGGAGGGGACGCGAGCGGTGTGCAGCCACACACAGCGCGGCCAGGCG 5296  
  
QY 7879 gccagggacccaacagagtagaccacgacctccagccactgctccccgaatgat 7938  
||||| : |||||  
Db 5297 GCCAGGGACCCAAAGCAGATGACCCAGCACCTCCACGCCACTGCTCCCCGGAATGAT 5356  
  
QY 7939 ttggaaccaaagtctaaactgagctcgagcccccgccctccctccgctcccatccc 7998  
||||| : |||||  
Db 5357 TTGGAACCAAGCTTAACAGCTCGCAGCCCCCGGCCCTCCCTCGCCCTCCCATCCC 5416  
  
QY 7999 gcttagcgtctggacagatggaacgagcgccctgtccagccccctagctcgctccgg 8058  
||||| : |||||  
Db 5417 GCTTAGCGCTGTGACAGATGAGCAGCGCCCTGTGCCAGCCCCCAGCTCGCTCGTTCGG 5476  
  
QY 8059 tccccacagactgccccagcaacagagattactgaaacccaagtcagccaggtggcg 8118  
||||| : |||||  
Db 5477 TCCCCACAGACTGCCCCAGCCAAACAGAGATTGCTGAAACCAAGTCAGGCCAGGTGGCGG 5536  
  
QY 8119 acaaaaggccaggtgctggcctggggggaacgagatgctccgagggactgactgttttt 8178  
||||| : |||||  
Db 5537 ACAAAAGGCCAGGTGCGCCTGGGGGAACGAGATGCTCCGAGGACTGGACTGTTTTT 5596  
  
QY 8179 cacacatcgttgcgcagcgttgggaagaaagggagatgtaaatgatgtgtggtttac 8238  
||||| : |||||  
Db 5597 CACACATCGTTGCCAGCGGTGGGAAGAAAGGAGAGATGTAATGATGTGTGTTGTTAC 5656  
  
QY 8239 aggttatatttgcatacttcaatgaataattcagatgttttaccgaagaaagactt 8298  
||||| : |||||  
Db 5657 AGGTATATTTTGATACCTTCAATGAATTAATTCAGATGTTTACGCAAGAGGACTT 5716  
  
QY 8299 acccagttactgctgtgtgttttgcattctcttaccgttcaagagcggtgtcgag 8358  
||||| : |||||  
Db 5717 ACCCAGTATTACTGCTGTGCTTTTGTATCTCTGCTTACCGTTCAAGAGGCGTGTGCG 5776  
  
QY 8359 gccagacgtcgtgaccccatcactcgaggacccaaggggggaggtgctgtcgacgc 8418  
||||| : |||||  
Db 5777 GCGCAGTCGTGTGACCCATCCTACGAGGACCAAGGGGGCGGGGATGCTGCTGTCAGCG 5836  
  
QY 8419 ccgctgtgtccctccctccctcccttcttggcgagaatgaattcgatcgctattctgtg 8478  
||||| : |||||  
Db 5837 CCCGCTGTGCTCCTCCTCCTCCTTCTTGGGAGAGATGAATTCGATTCGCTATCTGTG 5896  
  
QY 8479 gccgcattcgcaggggtgtgttattctgtcattacacacgctcgttctaataaaaa 8538  
||||| : |||||  
Db 5897 GCGGCATTTCGCGAGGCTGTGTGTTCTGCTGCTATTACACACGCTCGTCTTAATTAATA 5956  
  
QY 8539 gcgaattactcccaaaaaaataaaaaa 8571  
||||| : |||||  
Db 5957 GCGAATTACTCCAAAAAATAAAAAA 5989

RESULT 4

AF125671

LOCUS

DEFINITION Mus musculus silencing mediator of retinolic acid and thyroid

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source

gene

CDS

hormone receptor extended isoform (Smrte) mRNA, complete cds.  
AF125671 GI:4559295  
house mouse.  
Mus musculus  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 8544)  
Park, E.J., Schroen, D.J., Yang, M., Li, H., Li, L., and Chen, J.D.  
SMRTE, a silencing mediator for retinoid and thyroid hormone  
receptor corepressor  
Proc. Natl. Acad. Sci. U.S.A. 96 (7), 3519-3524 (1999)  
99199215  
2 (bases 1 to 8544)  
Chen, J.D.  
Direct Submission  
Submitted (03-FEB-1999) Pharmacology and Molecular Toxicology,  
University of Massachusetts Medical School, 55 Lake Avenue North,  
Worcester, MA 01655, USA  
Location/Qualifiers  
1..8544  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1..8544  
/gene="Smrte"  
160..7548  
/gene="Smrte"  
/function="transcriptional corepressor"  
/note="SMRte"  
/codon\_start=1  
/product="silencing mediator of retinoic acid and thyroid  
hormone receptor extended isoform"  
/protein\_id="AAD2972.1"  
/db\_xref="GI:4559296"

/translation="MSGSTQPVQAOVRAAEPRYPHGISTVQIARSHDTVDGLLEYOH  
HPRDITSLSPGSIQORRRPSTLSEFGSERSOELHLPESRTLPGLGKPDIEF  
TESKRRLLELLPDTLLRPSLLATGQSGSEDLTKDRLSLGKLEPVPSPSPHADPEL  
ELASRLSKELIONRLDRVDREITVQVQSLKLLKKQOOLEEAAAPPPEKVPSP  
PIESKLSLVOLIYDENRKAEEAARHLEGLGQVQLPQLNPDSRTORRKHENIKYQA  
MRKQPEIRKQELQRMQSRVQSGLSMSAAARSEHSEIIDLSELENLMQMKR  
LAVIRHVRDQQRKIFINMGLMDPMKVYKDRVTNMWSEQRDFREFMOPHN  
FGLTASLERKTVAECLVYVLTKNENKSLVRRYSRRKRSQOQOQOQOQOQOQ  
RSQSEKE  
TANSQGRKGRITRSMANEAHETATPQSSSELASMEMESSRWTBEEMETAKGLL  
EHNWSAIAIMVMGSKTVSOCKNFYFKKRNLDLILQHQHKLMEKERNARRKKKT  
PAAASEETAFPPAAEDEMEASGASANEELAEAEASQASGNEVPRVPGCSGPAVN  
NSDTEVPSPRSEATKDTGPKPTGTALPAATQPPVPPPEEPAAPAEPSVPDASG  
PPSPSPSPAAPPATVDKQOEAPAAPOTEDAKQEKSEAEIDVKGPEPEASEE  
PPSVKSDHKEETEPEEDKAKTEALVETVSEAPLKVEEAGSKAAVTKSGSGATODS  
DSSATCDADEVEPEGDKGLLSRPSLLTPAGDPRASTSPQPLDLKQLKQRAAAL  
PPIVTVHEPPREDTVPKPVPPPTQHLQPEGDSVQSGSGSGSGSGSGSGSGSGSG  
EAEKAPFPAPFPTEGQSYRLSPHAGHLRPLSHPPREVTKTSTRADPLFSTYPPGHLPL  
GLHDSARVLPFPPIINPPPLISSAKHPGLERQIGAIISQMSVQLRVPVSEHAKPMG  
PLMWELPVDKPKLGTALAPPVPEASPRASOYPCRRPOLRLYHPTPADVLYKGT  
LSRTVGEDSPSLDRADLTLPKGVHVEGKGVHLYEGMGVSYSCSGSGSGSGSG  
PHETAAPKRYDMMEGVGRVTSASIEGLMGRALPEOHSPLKEOHHSIGSTIQGLP  
RSVYAEQEDLIRREKLRREGTPPPPPPDLTETINRPLDPLGLPKLKPTEHGVV  
ATVKEAGSIETPREELRTPPELPAPRLKESITQGTPLKYDSGAPSGTGHKHDY  
RSITGSPRFPFALPLDINADARALERACEESLSRSGTSSGSGSITRGAAPVVP  
ELGKPROSLTYEDHGAFTSHLPRGSPVTTREPTRLQSGSLSSKASQDKRLTSTP  
RETAQSPHSTVPEHHPHISPYEHLRGVTDVLYRGIHPLAFOPTSTIPGILEAAA  
AAYVLPRLAPSTYPHYLPYLRGYDPAALENRQTLINDYITSOOMHNAASAMA  
QRADMLKGLSPRESSLAINYSAGPRGIIDLSOVPHLPVLPPTGCTATADRLAYLP  
TAPPPFSSRSHSSPLSPGPHTHLAKPATSSSERERERERERERERERERERERER  
WRPGEQSSGAGSSRPSATHQHSPTSPRTQDALQRPVSLNMTSKMGVTVSEPT  
PTVLRSTSTSPVRAAPPATFPATHCGTGLGTVLEGVPTLMPEVLLPKTSTRVAPERPR  
VDGHAFTLTPAREPASSPKSESLAPPSSSHTAIARTPAKSLAPHASDPDPC  
PTASDLHREKTSQKPSIOELRLSGLYHSGAGYSPDGVPIPSVSPSSPLTHDKGLS  
KPLELEKSHLEGLRHUKQKPKLSAEAAHLPLRPLPEOPSSPLLOAPGKIGH  
QRVYLAQHISEVITQDITRHHPPQOLSGPLPAPLYSPFGASCPLDLRPPSDLYLP



PDUGTARQSPHSEGGKSPBSKTSVLGSEDEATPEVPPPEGMTGPHARSAYVPLL  
YRDEGEPFMGSKPGNTSPQAPFSKLTESAMVSKKEINKLNTNHNDEEY  
NIGQETFPNIPATGACLTGTSQVQAEHASTNMGLEATIKALMGKYDDEEPP  
GLANAPLNASALPAAMPTTADGRSDHIALTSPGGGGKAKYKSPRSRKKAKSPAP  
GLASGDPSPSVSHSEGCNRRTPILNCSQYETSPAGSTFPPNPLINRLQAGVMA  
SPPPGLAAGSGPAGPHHAGDEEPPKLLCWQETLSDSE

	BASE COUNT	ORIGIN
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.000	0.000
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.000	0.000
16	0.000	0.000
17	0.000	0.000
18	0.000	0.000
19	0.000	0.000
20	0.000	0.000
21	0.000	0.000
22	0.000	0.000
23	0.000	0.000
24	0.000	0.000
25	0.000	0.000
26	0.000	0.000
27	0.000	0.000
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.000	0.000
32	0.000	0.000
33	0.000	0.000
34	0.000	0.000
35	0.000	0.000
36	0.000	0.000
37	0.000	0.000
38	0.000	0.000
39	0.000	0.000
40	0.000	0.000
41	0.000	0.000
42	0.000	0.000
43	0.000	0.000
44	0.000	0.000
45	0.000	0.000
46	0.000	0.000
47	0.000	0.000
48	0.000	0.000
49	0.000	0.000
50	0.000	0.000
51	0.000	0.000
52	0.000	0.000
53	0.000	0.000
54	0.000	0.000
55	0.000	0.000
56	0.000	0.000
57	0.000	0.000
58	0.000	0.000
59	0.000	0.000
60	0.000	0.000
61	0.000	0.000
62	0.000	0.000
63	0.000	0.000
64	0.000	0.000
65	0.000	0.000
66	0.000	0.000
67	0.000	0.000
68	0.000	0.000
69	0.000	0.000
70	0.000	0.000
71	0.000	0.000
72	0.000	0.000
73	0.000	0.000
74	0.000	0.000
75	0.000	0.000
76	0.000	0.000
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.000	0.000
83	0.000	0.000
84	0.000	0.000
85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.000	0.000
89	0.000	0.000
90	0.000	0.000
91	0.000	0.000
92	0.000	0.000
93	0.000	0.000
94	0.000	0.000
95	0.000	0.000
96	0.000	0.000
97	0.000	0.000
98	0.000	0.000
99	0.000	0.000
100	0.000	0.000

Query Match 58.8%; Score 5037.2; DB 94; Length 8544;  
Best Local Similarity 78.3%; Pred. No. 0;  
Matches 6770; Conservative 0; Mismatches 1493; Indels 387;

Qy	1	catgtcgggtccacacagatttgtgcacagactgtgagggccactgagcccgctacc	60
Db	159	CATGTAGGATCCACACAGCCTGTGGCACAGATATGCGGGTGTCTGAGCCCGCTACCC	118
Qy	61	gccccacagcctttaccacagtgcagatgcccgagcgacacagagcgtgggctct	120
Db	219	ACCCCATGGCATCTCTACCCGGTGCAGATAGCCCGGTCCACACAGCAGTGGGGTGT	278
Qy	121	ggagtaccagcacactcccgactatgctctccacctgtgcgggtgctccatcatcca	180
Db	279	TGAGTACCAACACACCCCGTGACTACACCTCACACCTGTACCCTGGTTCATCATCCA	338
Qy	181	gccccagggcgaggccctcctgtgtgtgatgtccagcccggaatgaacggtccca	240
Db	339	GCCACAGAGGAGCGGGCTTCACTGTGTCTCAGAGTTCCAGCCTGGGAGTGAACGGTCTCA	398
Qy	241	ggagctccacctggggccagagtcacactcaatacctgcccgagctggggaagtcaagat	300
Db	399	GGAGTCTCAACTGTGGCCCTGAGTCCCGCAGTTCTCTGCTGAGCTGGGAAGCCGACAT	458
Qy	301	ggagttcattgaagcaagcgccctcgctgagctgtgctgtgctgtgacccctctcgacc	360
Db	459	AGANTTCAACGAGGACGAGCGCCCGCTTGAGACTACTACCCGATACCTGTCTGCGCCC	518
Qy	361	gtcaccctctgtggccacggggccagcctgggagtctgaagacctcaccaaaggaccgtag	420
Db	519	ATCACCCCTGTGTGCCACTGGGAGCGGAGTGGGTCTGAAGACCTTTACCAAGGACCGTAG	578
Qy	421	cctgaacgggcaagctggaaacgggtgtctccccccagccccccgcacactgacctgagct	480
Db	579	CCTGGCAGGCAAGCTGGAGCCTGTGTCACTCCCTCAGTCCCGCCGACGCTGACCTGAGCT	638
Qy	481	ggagctgtgtccgcacggctgtccaaagagagctgtatccagaaca---tgagacgct	537
Db	639	AGAGCTGGCCCATCTCGACTGTCCAAGGAGGAGCTGATCCGAACAGANTTGGACCGCT	698
Qy	538	ggaccgagagatcaccatggtagagcagcagatctcttaagctgaagaagaagcagcaaca	597
Db	699	GGACCGTGAGATCACCATGTGTAGCAGCAGAGATCTCCAAGCTGAAGAAGACGACGACAA	758
Qy	598	gctggaggagggtgtccaaagccgcccagcctgagaagccggtgtcaccccgccgccc	657
Db	759	GTTGGAGGAGGAGCGCCCAAGCGCGCCGAACCCGAGAAAGCGCTGTCTGCCACACCCCAT	818
Qy	658	cgagtccaagcaccgacgctgtgcagatcatctacgacgagaaacccggaagaaggctga	717
Db	819	AGAATCAAAAGCACCGAAGCCTGTGTCCAGATCATCTACGATGAGAACCGGAAGAAAGCCGA	878
Qy	718	agctgcacatcggattctggaaggcctggggcccagggtggagctgcgctgtacaacca	777
Db	879	AGCGGCACACGGATCCTAGAAGGCCTGGGGCCCCAGTGGAGTGCCTTCTGTACAAACCA	938
Qy	778	gcoctcgcacacccggcagtatcatgagaacatcaaaataaaacccagcgcatcggaagaa	837
Db	939	GCGCTGTGACACACGCCAGTACCATGAAACATCAAAATAAACACCGCGCATGCGGAAGAA	998
Qy	838	gctaaatttgtactcaagaggagggaatcacgctcgggaacaatatggaagcagaagtctgt	897
Db	999	GCTGATCTTGACTTTTAAGCGGAGGAACACCGCGCAAGCAGTGGGAACAGCGTCTTG	1057

[illegible]







Db 4129 GTGGAGGGCGCAGGAGACTACTTACGGCGGGAGGCCAAGCTCTTGAACGAGAAAGGACA 4188  
Qy 4151 cctcgcgcacccacccctcagcgagcctgacggagcctcacaagacgagcgc----- 4204  
Db 4189 CCACCTCCCCACGACACCTCGGGACCTGACTGAGACCTACAAGCCCGCGCCCTGGAC 4248  
Qy 4205 ---ctgggccccctgaagctgaagcggcccatgagggccctggtggccacggtgaagag 4261  
Db 4249 COTCTGGTTCCTTGAAGCTGAAGCCGACTCAGGAGGTTGGTGAAGTGAAGAG 4308  
Qy 4262 gggggccctccatccatgaatcccgccgagagctgcggcacacgcccagcgtgcc 4321  
Db 4309 GGGGCCCTCTATCCATGAGATCCCGAGAGAGAGCTGCGCGCACCTGAGTATACC 4368  
Qy 4322 ctggccccgcgcctcaagagggctccatcacgagggcaccgcgctcaagtacgac 4381  
Db 4369 CTGGCACACGCGCTCTGAAGGAGGTTCCATCACCCAGGGCACCCCACTCAAGTACGAC 4428  
Qy 4382 accggcgctccacactggctccaaaaagcacgagctacgctccctcatcgccgc 4441  
Db 4429 TCTGGGGCACCTCTCCACTGGCACCAAGAAACACGACGTCGCGTCCATCATCGGCAAGCCCC 4488  
Qy 4442 ggcggagcttccacaccgctcacccgctgcatgtgatggcgagcgcgggacactgaa 4501  
Db 4489 GCGCGGCTTTCCCTGGCTGCACCCGCTGGACATAATGGCTGAGCGCGGACACTGGAG 4548  
Qy 4502 cgtgctgtacgagagagcctgaagcgcgccgagggacggcagcagctcggggggc 4561  
Db 4549 CGTGCTGCTATGAGAGAGTCTGAAGAGCCGGTCAGGGACCAAGCTGGTCAGGGGGC 4608  
Qy 4562 tccattgcgggcgccccggtgattgctgagctgggtgaagcggcgagagcccc 4621  
Db 4609 TCCATCACACCTGGGGCTCCAGTCTGCTGCTGAATGGGCAAGCCACGGCAAGGCCA 4668  
Qy 4622 ctgacctatgaggaacacggggcaccctttgcccgcacctccacagaggttgcgcctg 4681  
Db 4669 CTGACTTACGAAGACACGGGGCACCTTCACCACTAGCTGCCAGTGGCTTCCCTGTG 4728  
Qy 4682 accatcgggagccacgcgcgctcagagggcagccttctgctcagcaagcatcc 4741  
Db 4729 ACCAGGAGGACCCACGCCAGCGCTTCAGGAGCGACGCTCCATCCAGCAAGCGCTCC 4788  
Qy 4742 caggaccgaagctgaagctgacgcgctcgtgagatcgccaagtcgccacagcaccgtg 4801  
Db 4789 CAGGACCGGAAGCTGACATCTACACCCCGGAGATCGCAAGTCCCCACACAGCACTGTG 4848  
Qy 4802 cccgagacacccacacccatctcgccctatgagacatgcttcggggcggtgagtg 4861  
Db 4849 CCGGAGCACCCCTCACCCCATCTCCCTCCATGAGCACTTGTCTCCGGGGCGTGACTGGT 4908  
Qy 4862 gtggacctgtatcgagcacaatccccctggccttcgaccaccaactccatcccccggc 4921  
Db 4909 GTGGACCTGTACCGTGGTGCATCCCATTTGGCTTTGACCCCACTCCATACCCCGAGGG 4968  
Qy 4922 atccctctgga---cgcagccgctgctactactctgccccgacacactggcccccaacccc 4978  
Db 4969 ATCCCTCTGGAAGCAGCAGCGCAGCGCTACTACCTGCCCCCGCACTTGGCCCCAGCCCC 5028  
Qy 4979 acctaccgacactgtaccaaccctacatcgcggcgtaccacccgacacggcgcgctg 5038  
Db 5029 ACCTACCCACACCTGTATCCCACTTACCTCATTCGCGGCTACCTGTACGACGGCGCCCTG 5088  
Qy 5039 gagaacggcgagacatcatcaatgaactacatcactcgagcagatgacacaaacag 5098  
Db 5089 GAGAACCCAGACCATCATCAATGACTATACCTCGAGCAGATGCACCAACACCT 5148  
Qy 5099 gccaccgcatggccccagcagctgatgctgaggggctctcgcgcgcgcgagctcctcg 5158  
Db 5149 GCTTCGGCATGGCCAGCGTGTGACATGCTGAGGGGTCTGTACACGCGAGAGTCTCTCG 5208  
Qy 5159 ctggcactcaactacgctgcggttccccgaggaatcatcgacctgttcccaagtgccac 5218  
Db 5209 CTGGCCCTCAATTATTCGGTGGGCCAAGAGAGGCATTATCGACCTGTCCCAAGTGCACAC 5268

Qy 5219 ctgacctgtgctgccccccgacacccagcagccacccagccatggagcccttgc 5278  
Db 5269 CTGCCCGTGTGGTGCACCAACGCGCAGGACCCCTGCGACCCGATCGACCGCTTGC 5328  
Qy 5279 tactccccacgcgcgcaccccttcagcagcgcgcacagcagctcccccactcccca 5338  
Db 5329 TACCTCCCACTGTGCGCCCACTTTCAGCAGCGCCACAGTAGCTACCGGTGCCCCA 5388  
Qy 5339 ggaagtccaacacacttgacaaaacacacccacacgctcctcgtccgagcggagcagac 5398  
Db 5389 GGAGCGCCCACTCACTAGCTTAACCAACTGCCACATCTTTCATCGGAGGGGAACG---- 5444  
Qy 5399 cgggctcagagcgggacccgggacgagcggggaagtgccatcctcagctccaccag 5458  
Db 5445 -----GGAACGTGAGCGGAACGAGACAAGTCCATCTCTCAGCTTACCAC 5490  
Qy 5459 acggtgagcagcaccacatctggagacgctgtacagagcagcagcagcagcggc 5518  
Db 5491 ACAGTGGAGCATGCACTTGGAGACTTGGTACGAGCAGAGCAGCGGGCTGGGGC 5550  
Qy 5519 agcagcggcgggggtgggggagcagcgcgcgcgcctccactcccatgccaccag 5578  
Db 5551 AGACCGG-----CCCCGCTCCACACCCACCACG 5580  
Qy 5579 cactgcacatctccctcctggaccagcagctccctccagcagacccagctgtgttcc 5638  
Db 5581 CACTCGCCCATCTCCCGCGGACCCAGGACGCTTGCAGCAGAGGCCAGTGTGCTGCAC 5640  
Qy 5639 aacacagcagtgaaaggttatcatcaccgctgtggagccacgcaagccccagctcctg 5698  
Db 5641 AACACGAGCATGAAGGGGTGGTCACTCCGTGGAAACCGCGCACGCCACCGTCTGAGG 5700  
Qy 5699 tccactccactcctcactcctgcagcagctgcacattccacactccacccactgc 5758  
Db 5701 TCCACTCTCACCTCTCGCTGTGCGCCAGCTGCCACATTCCTCCACTGCCACCACCTGC 5760  
Qy 5759 ccaactggcgggcaccctcgatgggtctaccctacccctcagcagccgcttctgtgcc 5818  
Db 5761 CCACTTGGTGGCACCTTGAAGGGGTCTACCTACCTCATCGGAGCCGCTCTGTATACC 5820  
Qy 5819 aaggaggcccccggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgccttc 5878  
Db 5821 AAGGAGACTCTCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 5880  
Qy 5879 ctgcgaagccccagc 5938  
Db 5881 CTCACCAACCCCGCGCGCG-----GGAGCGCGCTCTCTACCCAGCAAGAGCTCC 5931  
Qy 5939 gagccccggccctagtgctcctgtctcgtgcacgcgcacccatcgcgcgcgcgcgcgcgc 5998  
Db 5932 GAGCCCCGATCCCTAGCACCCCGCGAGCTCCAGCCACACAGCCATCGCGCGCACCCAGCA 5991  
Qy 5999 aagaacctcgcacatcacaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcctg 6058  
Db 5992 AAGAGCTTGGACCCACCATGCGAGTCGGAGCGCGCGCGCGCGCGCGCGCGCGCTCA 6048  
Qy 6059 gaccgcacggggaagactcaagtaaaccttttccatccaggaactggaactcgt 6118  
Db 6049 GATTCGACCCGAGAAAGACTCAAAAGTAACCTTTTCATCCAGGAATTTGAACCTCGT 6108  
Qy 6119 tctctgggtttaccac---ggcagcagctacaccccggaagggtgagccgctcagccct 6175  
Db 6109 TCTCTGGGTACCACAGTGGAGCTGGCTACAGCCCGCGATGGGTGGGTGAGCCCCATCAGCCG 6168  
Qy 6176 gtgagctcaccagctctgaccccaagaaggggtcccccaagcactcctggaagagctcag 6235  
Db 6169 GTGAGCTCCCGCCTGACCCACGACAAAGGGGTCTCTCCAAACCTCTCTGGAAGAGTAGAG 6228  
Qy 6236 aagagcaactcggagggagctgcggcccaagcagcagcccgctgaagcttggcggg 6295  
Db 6229 AAGAGCCACTTGGAAAGGGAGCTGCGGCACAGCAGCGCGGCCCATGAAGCTCAGCGCG 6288











QY 661 gtccgaagaccacccagcctggtgcaagatcatctacagcagagaaccgggaagggctgaagc 720  
DB 1285 ATCAAGACCCGAAGCCCTGGTCCAGATCATCTACGATGAGAACCAGCGGAAGAAAGC 1344  
QY 721 tgcacatcgattctggaagcctggggcccccaggtgagctgagctgctgtaacaccagcc 780  
DB 1345 CCACACCGGATTCCTAGAAAGCCCTGGGGCCCCAGGTGAGCTGCCTCTGTCAACACGCC 1404  
QY 781 ctcccgaacccggcagttatcatgagaacatcaaaataaacacagcgcatcggaagaagct 840  
DB 1405 GTCCTGACACGCCAGTACCATGAATAATCAATAAACACGCGGATGCGGAAGAACT 1464  
QY 841 aatctgtacttcaagagaggaatacaagctcgggaacaaatggaagcagaagttctgcca 900  
DB 1465 GATCTTGTACTTAAAGCGGAGGAACACCGCGCAGCAGTGGGAACACAGCGCTTCTGCCA 1524  
QY 901 ggcgtatgacacagctcatggaagcccttggaataaaaggtgagcgcatcgaaacaaacc 960  
DB 1525 GCGCTATGACCAAGCTCATGGAGGCCGTGGGAGAAAGGTAGAGCGCATAGAGAACTATCC 1584  
QY 961 ggcgcgagcggcgaagagagcaaggtgcgagttactacgaaagcagttccctgagat 1020  
DB 1585 GCGAAGGAGGCCAAGGAGACCAAGGTGAGGAGTACTACGAAACAACTTCCCGGAGAT 1644  
QY 1021 ccgcaagcagcgagctgcaggagcgcatgcagagcaggggtggccagcgggcgagttgg 1080  
DB 1645 CCGCAAGCAGCGGAGCTGCAGAGCGCATGCAGAGCAGGCTGGCCACAGCGTGGCAGTGG 1704  
QY 1081 gctgtccatgtggccgcccagcagcagcagcagcaggtgtcaagatcatcgatggcctctc 1140  
DB 1705 GCTCTCCATGTGGGTGGCCCGCAGTGAGCATGAGTTTCTGAGATCATGTGTTGGTGTCT 1764  
QY 1141 agagcagagaaacctgggaagcagatgcgcagctgcccgtgaccccgcccatgctata 1200  
DB 1765 TGAGCAGAGAACTTGGAGAACGATGCGCAGCTGCCGTGATCCCGCCCATCTTTGTA 1824  
QY 1201 cgaagctaccagcagcgcatcaagttcatcaacatgaacgggcttatggccgaccccat 1260  
DB 1825 CGACCGGACCGCAGCAGAGGATCAAGTTTCAATCAATGAATGGACTCATGGATGACCCCAT 1884  
QY 1261 gaaggtgtacaaagaccgcccaggtcatgaaatgtgagtgagcagaggaaggaacatt 1320  
DB 1885 GAAAGTCTACAAAGGACCTCAGGTTTACCAATGTGGAGCAGCAGGAGGAGGACACCTT 1944  
QY 1321 ccgggagaagttcatgcagcatcccaagaactttggcctgcatcgatcatcttcctggagag 1380  
DB 1945 CCGTGAGAAGTTATGACGACACCTTAAGACTTTGGCCTGATGGCTCATTTCTCGGAGAG 2004  
QY 1381 gaagacagtggtgagtcgctctctctatactactcgtactaagaagaatgagaactataa 1440  
DB 2005 AAAGACGGTCGTGAGTGTCTCTATTACTTACCTGACCAAGAAATGAAAATTACAA 2064  
QY 1441 gagcctgtgagcagagctatcgccgcccggcgaagagccagcagcaacaacagcagca 1500  
DB 2065 GAGCTTGTGAGGGAGCTATCGCGCCGTGGCAAGAGCCAGCAGCAGCAGCAGCAGCA 2124  
QY 1501 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1560  
DB 2125 ACAACAGCAGCAGCAGCAGCAG-----ATGCCAGGAGCAGCAGGAGGAGAA 2172  
QY 1561 agatgagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1620  
DB 2173 GGAGGAGAAGGAGAAGGAGGAGGCGCGCAAGGAGGAAGCAAGGATGCGGAGAA 2232  
QY 1621 cgaagaagaacctctcaagagagaagacagacacctcaggggaggaagaacagcagca 1680  
DB 2233 CGAGAGGAAGAACTTCACAAAGGAGAAGAGGCGCGCAAGGAGGAAGCAAGGATGCGGAGAA 2292  
QY 1681 gaaggaagtgctggtcctccaaagcgcaaaactgccaaacagcaggaaggaagcgaag 1740  
DB 2293 GAAAGAGGCCCTGGCCCTCCAAAGGCCGCAAACTGCCAAACAGCAAGGCCCGCCCAAGG 2352  
QY 1741 ccgcataccccgctcaatggttaatgaggcccaacagcagagagagccatcaacccccagca 1800

DB 2353 CCGTATCATCGCGCTCCATGGCCACAGAGCCAAACCATGAGGAGACAGGCCACCCACAGCA 2412  
QY 1801 gagcgccagctgctgctccatgagagctgaaatgagagttctcgctgagcagaagaagaat 1860  
DB 2413 AAGTTTCAGAGCTGGCTTCATGAGAGTGAAGCAGAGAGTTCTCGCTGGACTGAGGAAGAT 2472  
QY 1861 ggaacacgcaagaaaggtctcctggaacagcgccgcaactggtcgccatcgcccgat 1920  
DB 2473 GGAGACACAAGAAGAGCCCTCTGGAACATGGGAGGAACCTGGTCAGCATTTGCCGAT 2532  
QY 1921 ggtgggtccaagactgtgctgagctgtaagaactctacttcaactacaagaagga 1980  
DB 2533 GGTGGGCTCCAAGACCGGTCTCCAGTGTAAAGACTTCTACTTCAACTACAAGAAGAGCA 2592  
QY 1981 gaacctcgatgagatcttgcagcagcacaagctgaagatgagaaggaaggaagcgcg 2040  
DB 2593 GAACCTGGACCAAAATCTTTCAGCAGCACAACTAAGATGGAGAAGAGAGAGCACTCG 2652  
QY 2041 gagaaagaagaagaaagcgccgagcgccagcgagaggtgcatcccgccggtggt 2100  
DB 2653 GAGGAAGAAGAAGAACCCAGCTGCGCGCAGGAGGAGACAGCTTCCACCTTCCGCGC 2712  
QY 2101 ggaagatgagagatggaagcggtcggtgagcggaatgagggagagatggtgagga 2160  
DB 2713 TGAGSACCAAGAGATGGAAGCATCAGGCGCAAGTGCCAAATGAGGAAGAGCTGGCGGAGA 2772  
QY 2161 ggcgtgaacctatactgctctggaatgaggtgcccaga---ggggaaatcgagtgccc 2217  
DB 2773 GGCAGAACCTTCACAGGCTCTGGGAATGAGTTTCCAGAGTTGGGGAGTGCAGTGGCCC 2832  
QY 2218 agcactgtcaacaacagctcagacacgagagatccccctctcctacactgagcgcc 2277  
DB 2833 AGCTGCTGTCAACAAACACTCTGATCTGAGAGTGTCCATCCCGCCCTTCAGAGCCAT 2892  
QY 2278 caagacacagggcagaatgggccaagccccagccacctgggcgcgagcgccacc 2337  
DB 2893 GAAGSACAC-----TGGGCTTAAACCCACTGGCACATGAAGCATTTGCCGCTGCCAC 2943  
QY 2338 cccagggcccccacccccacacagagagacatccccgggcccccatgagaccacccccgc 2397  
DB 2944 CCAGCCACTCTTCTCTCCAGAAACCCGCGAGTAGCCCTTGTGAGCCCTCCCCAGT 3003  
QY 2398 ctctgaagcccccagagccccctagccccccagcagcccccatcgccctctgacatcc 2457  
DB 3004 CCTGTATGCCAGTGG-----CCACCATCCCCAGAGCCTTCCCATCACCCTGCCGACCCCC 3059  
QY 2458 tctctgtgtcccccaagggaggaagggagagagcagcagcagccccagtgagga 2517  
DB 3060 GCCTACTGTGCACAAAGGATGAACAAGAGCCCGCGCTGCTCCAGCTCCCCACAGAGAGA 3119  
QY 2518 gggggagagcagaagcccccccgcggtgag-----gagctggcag 2558  
DB 3120 TGCCAAAGAGCAGAAAGTGTGAGGCCGAGGAGATCATGTGGGAAAGCCAGAGAGCCCG 3179  
QY 2559 tggacacaggaagcgagagagccctcaagcagcagtcacaggaaggaagccagagag 2618  
DB 3180 AGGCTCTGAGAGAGCCCCCGGAGAGTGTAAAGAGTGACCACAAGGAGGAGACCCAGGAG 3239  
QY 2619 gggcggc---caagggcaagggagcggtgagggccagcgccagggggggcgctca 2675  
DB 3240 AGCCTGAGACAAAAGCCAAAGGCGCACAGAGGCCATTGAACCTGTGTCTGAGGACACACTTA 3299  
QY 2676 aggcagagaagaagggcgagcgagcgagggccacactgccaagagctcgggcgcc 2735  
DB 3300 AG---GTGGAGAGGCTGTAGCAGGAGCGCTGTGACCAAGGGTTCCAGCTCAGTGTGCCA 3356  
QY 2736 cccaggaagcagctccagttctacctgagtcagagcagagtgagagcagggcg 2795  
DB 3357 CCCAGGACAGTACTTCAGTGCCACCTGCAGTGCAGTGCAGTGCAGTGCAGCAACCCGAGGAG 3416  
QY 2796 gcgacaagaacccgctgctgttcccccaagggccagcctctcactccccactgagccccc 2855



Db 3417 CTGACAAAGGCGAGGCTGCTGTGTACCAAGGCCAGCCTCTCTACCCCGGTGGAGATCCCC 3476  
Qy 2856 gggccaatgctcaccacagcaagccactggaccctgaacagctgaagcagcagcagcgtg 2915  
Db 3477 GGGCCAGTACCTCGCCCGCAGAAAGCGCTGGACCTGAAGCAGCTGAAGCAGGAGCAGCGG 3536  
Qy 2916 ccatcccccaccatccaggtaccacaaagtccatgagccccccgggagcagcagctccca 2975  
Db 3537 CCATCCCCCTATCCAGTCCACCAAGGTCATGAGCCCGCCCGGGAGGACAGTACCCC 3596  
Qy 2976 ccaagcagctcccccacccacccacccacccacccacccacccacccacccacccaccc 3035  
Db 3597 CAAGCCAGTTCCCGCTGCTGCTCCACCCAGCAGCAGCTACAGCCAGAGGGTGTGCT 3656  
Qy 3036 ctacagcagctggcagcagccccggggcgaagagcagcagcccccccccagccagaca 3095  
Db 3657 CTCAGCAGTCGGGAGGAGTCCAGTGGCAAGTCCCGAGGCCAGTGTCTCTGCCGAGA 3716  
Qy 3096 agga-----ggccttcgagccgagggcccaagaagctgcgtg 3131  
Db 3717 AAGAGGACAGAAACCCGCATTCTTTTCGGGCTTTCCCAACTGAGGGCCCAAAGCTACCGA 3776  
Qy 3132 gggaccccccttgctgagcttcggcctgccccttcctcccgctgcccccccgtagtgatca 3191  
Db 3777 CTGAGCCCCACGCTGTGTCTATCGGGCCTTGCCCTTCTCCCATCCCTCCACGGGAGGTGATCA 3836  
Qy 3192 aggcctcccccagcagccccgaacccctcagccttcctcagctccacactgtaacccac 3251  
Db 3837 AGACTTCCCAACAGCCGCTGACCCCTTGCCCTTCTCTACACACCCCGGCTACCCGCG 3896  
Qy 3252 tgcctcctggcctccatgacactgccccggcctgctcctccgcccacccacccactctcca 3311  
Db 3897 TGCCTCTGGGCTCCACGATAGTGGCCGCGCGCTCTCTGCGACG---TCCCCCATCTCTA 3953  
Qy 3312 acccgctccctcctatctcctctgccaagacccccacagcgtcctcgagaggcaaatagtg 3371  
Db 3954 ACCCCCAACCCCTCATCTCCCTGCAAGATGTCGGGGCTACTTGAGAGGAGCAGCTGGGTG 4013  
Qy 3372 ccatctcc---caaggaatgctggtccagctccagctcccttactcagcagcagcagcag 3428  
Db 4014 CCATCTCCACGAGGGGATGTCACTGACAGCTTGTGTGCTCTCACTCAGAGCATGCCAAG 4073  
Qy 3429 ccccggtggccctgtcaccatggggtgcccctgcccctgaccccccaaaagctggcac 3488  
Db 4074 CCCCCTATGGGCCCTCTACCATGGGCTTGCCCTTGCCGTGGACCTTAAGAAGCTG---- 4129  
Qy 3489 ccttcagcggagtgaagcagagcagctgtcccccaggggcccaggtggccacccggaga 3548  
Db 4130 ----- 4129  
Qy 3549 gcttggggtgtccacacagccagaggcgctcgtgtagagggacagctctgggctcag 3608  
Db 4130 -----GGGACAGCATGGGCTCGG 4148  
Qy 3609 ttccggggcgaagcattaccacaaaggcattccacagacacgggtgcccctcgacagcaca 3668  
Db 4149 CCACAGTGGGAAGCATCACCAAGGCGCTCCCGCAGTACCCGGG-----CTGACAGCGGCC 4202  
Qy 3669 tcacataccggcgttccatcaccacacggcagcagcagctgacgtctctgtacaaaggcaca 3728  
Db 4203 CCAGCTCAGAGGCTCTATACCCACGCGACGCGCGCAGAGGTCTCTTACAAAGGTATCA 4262  
Qy 3729 tcaccaggtatcagcagagcagccagctgcttgacccggcggcggcggagacagcc 3788  
Db 4263 TCAGGAGGATGCTGGTGAGGACAGCCCAAGTGGCTTTGACCGGGCAGCAGGAGACCCC 4322  
Qy 3789 tggccaaggccagctcattacgaaggaagggccacagctcttctcctatgaggtg 3848  
Db 4323 TGCCCAAGGGCCATGTATCTATGAGGCGCAAGAAAGGCCAGCTCTATCTATGAGGTG 4382  
Qy 3849 gcatgtctgacccagctgtcccaaggaggcagcagcagctcagacccccccatg 3908  
Db 4383 GTATGTCGCTCAGTGTCTAAGGAGGATGGAAGGAGCAGCTTCGGGGCCCAACCCCATG 4442

Qy 3909 agagccgcgcaccccaagcgacacctatgacatgtagagggccgctgggcaagcactct 3968  
Db 4443 AGACTGCCCGCCCTTAACACGACCTATGACATGATGAGGGCCGTGTAGCGAGGACTGTCA 4502  
Qy 3969 cctcagcagcagcaggaaggtctcattgggcccgtgcatcccgccgagcagcagcagcccc 4028  
Db 4503 CCTCAGCCAGCATAGAGGAGCTCATGGCGCGGCCATCCCTGAGCAGCA-----CAGCC 4556  
Qy 4029 accactcaagagcagcagcaccacatcccggttcacacacaaagggatccctcggtct 4088  
Db 4557 CCCACTCAAGGAGAGCATCACATCCGAGGCTCCATCAGCAAGGCATCCCGAGGTCTCT 4616  
Qy 4089 acgtgagggcacagagagactacctcgtcgggagggcacaagctcctaaagcgggagggca 4148  
Db 4617 ATGTGAGGCGCAGAGGAGCTACTTACGGCGGAGGCCCAAGCTCTTGAAGCGAGAGGGA 4676  
Qy 4149 cgcctccgccccccacccgcctcacgggacctgacgagggcctacaaagcagcagggcc--- 4204  
Db 4677 CACCACCCACCCACACACCTCGGGACCTGACTGAGACCTTACAAGCCCGCCCTGG 4736  
Qy 4205 -----ctgggccccctgaagctgaagccggccccatgagggcctggtggccacgggtgaagg 4259  
Db 4737 ACCCTCTGGGTCCCTGAAGCTGAAGCCGACTCAAGAGGGTGTGGTAGCACTGTGAAG 4796  
Qy 4260 agggggcgccgtccatccatgagatcccgcgagagagctcgccgacacgcccagctgc 4319  
Db 4797 AGCGGGCGCTCTATCATGATGATCCCGAGAGAGAGCTGCGCGCGACACCTGAGCTAC 4856  
Qy 4320 ccttgcccgcggcgccgtcaaggaggctccatcacgagggcaccgccctcgaagtacg 4379  
Db 4857 CCTGGCACACCGGCTCTGAAGAGGGTTCCATCACCCAGGGCACCCACCTCAAGTAGC 4916  
Qy 4380 acacggcgcgctccacacactggctccaaaagcagcagtaagctccctcactcgccagcc 4439  
Db 4917 ACTCTGGGGCACCTTCCACTGCGACCAAGAAACACGAGCTGCGCTCCATCATCTCGCAGCC 4976  
Qy 4440 cggcgcgagctgtccacacccgtgaccccgctggtatgtagcccgacgcccggcactgg 4499  
Db 4977 CCGCGCGGCTTTCCCTGCGCTGCACCGCTGGACATATATGGCTGACGCCCGGCGACATGG 5036  
Qy 4500 aacgtcctgtctacgagagagcctgaagcggccgggaccgccaagcagctcggggg 4559  
Db 5037 AGCGTGCCTGTATGAAGAGAGTCTGAAGAGCGCGTTCAGGGACACGAGCTGTGCGAGGG 5096  
Qy 4560 gctccatgcgcgcgcccgcccgctcattgtcctgagctggttaagcggcgagagacc 4619  
Db 5097 GCTCCATCACAGTGGGCTCCAGTCTGCTGCTTGAACCTGGGCAAGCCACGGCAAGGCC 5156  
Qy 4620 ccttgacctatgaggaaccacggcgccacctttgcccggccacctccacgaggttcgccc 4679  
Db 5157 CACTGACTTACGAAGACACCGGGGACCCCTTACACAGTCACTGCGCAGCTGTCTCCCTG 5216  
Qy 4680 tgacctgaggagccccacgcgcgctgagggagggcagccttctgtccagcaaggcat 4739  
Db 5217 TGACCACGAGGAGGCCACGCCACGCCCTTCAGAAAGGAGCGCTCTATCCAGCAAGGCT 5276  
Qy 4740 cccaggaacgaagctgacgtgacgctctgagatgcgaagtcccccgacacagccag 4799  
Db 5277 CCCAGGACCGGAAGTGCATCTACACCCCGGAGATGCGCAAGTCCCGACACAGCAGCTG 5336  
Qy 4800 tggcgagcacaccacacccccctcgtccctatgagcactgctcggggcgtgagtg 4859  
Db 5337 TGCCGAGACACACCCCTCACCCCATCTCCCCATGAGCAGCTTGTCTCGGGGCGGTGACTG 5396  
Qy 4860 gcttgagcctgtatcgagccacatccccctggccttcgacccccccctccatcccccgcg 4919  
Db 5397 GTGTGAGCTGTACCGTGTGTACATCCCATTTGGCCTTTGACCCCACTTCCATACCCCGAG 5456  
Qy 4920 gcatcctcttggga----cgagccgctgctactactcctgcccgcagacactgcccccaacc 4976  
Db 5457 GGATCCCTCTGGAAGCAGCAGCGCGAGCCTACTACTGCGCCCGGCACTTGGGCCCCAGCC 5516



QY 4977 ccactaccgcgaactgtacccacacctaccctacatcccggtctaccccgagacacgagcggcg 5036  
Db 5517 CCACCTCCACACCTGTACCCACCTTACCTCATCCGCGGTACCTGACACGGCGGCC 5576  
QY 5037 tggagaacgagcagacatcatcaatgactacatcaatcgcagagatgacacacaaca 5096  
Db 5577 TGGAGAAGCGCAGACCATCATCAATGACTACTCATCACCTCGCAGCAGATGCACCAACG 5636  
QY 5097 cggcaccgcatgcccagcagactatgctgaggggctctcgcccagagctct 5156  
Db 5637 CTGCTCCGCCATGCGCCAGCGTGTGACATGCTGAGGGGTCTGTACCGGAGAGTCC 5696  
QY 5157 cgtggaactcaactacgtcggggtcccaggggcatcatgacacgtgtcccaagtgcac 5216  
Db 5697 CGTGGCCCTCAATTTATGCGCTGSCCCCAAGAGGCATTTATCGACCTGTCCCAAGTGCCAC 5756  
QY 5217 acctgctgtgctgtgcccgcgacacacagcgcacccccagcgcgcgcgcgccttg 5276  
Db 5757 ACCTGCCGTGTGTGTGCCACCAAGCGCAGGACCCCTGCCACCGCCATCGACCGCTTG 5816  
QY 5277 cctactccccacgcgcgcgcgcgccttcagcagccacagcagctccccactctccc 5336  
Db 5817 CCTACTCCCCACTGCGGCCCCACCTTCAGAGCGGCCACAGTAGCTACCGCTGTCCC 5876  
QY 5337 caggaggtcccaacacacttgacaaaacacacacacacacacacacacacacacac 5396  
Db 5877 CAGGAGGCCCTCACCTAGCTAAACCACTGCCACATCTTCATCGGAGCGG----- 5929  
QY 5397 accggggtcgagagcggaacgagcggtgagcgagcggaagaaagtccactcactcaca 5456  
Db 5930 -----GACGGGAACGTGAGCGGGAACGAGACAAAGTCCATCTCATCGTATCCA 5978  
QY 5457 cgacggtgagcagcaccatcttgagacctggtacagcagcagcagcagcagcagc 5516  
Db 5979 CTACAGTGGAGCATGCACCCATCTGAGAACCTGTGTACGAGCAGAGCAGCGGGCTGGG 6038  
QY 5517 gacgagcgcggggtggtgggagcagcagcgccccgcgcctccactcccatgcccacc 5576  
Db 6039 GCAGCAGCGGC-----CGGCTTCCACACCCACC 6068  
QY 5577 agcactgcccactctccctcggaacacagatgcctccagcagagacacagctgtcttc 5636  
Db 6069 AGCCTCGCCCATCTCCCGGACCCAGGACCGCTTCGACGAGAGGCGCCAGTGTGCTGC 6128  
QY 5637 acaacacaggtatgaaggtatcatcacctgttgagcccgacgaagcccagctctctg- 5695  
Db 6129 ACAACACGAGCATGAAGGCGTGTGTCCTCGTGTGAACCGCGCACGCCAGGTCTGA 6188  
QY 5696 -----aggtcaactccactctcactcactcactcactcactcactcactcactc 5747  
Db 6189 GGTGGGCGAGGTCCACCTCCACCTCTTCGCTGTCCGCCACAGCTGCCACATTCGCCACTG 6248  
QY 5748 ccaccactgcccactggcgccactctcgatggggtctaccctaccctcatgagcccg 5807  
Db 6249 CCACCACTGCCACTTGTGTGGACCTTTGAAGGGGTCTACCTTACCTCATGTGAGCCCG 6308  
QY 5808 tcttgctcccaagagggcccccggtgccccggcagagcgcccccgagcagacacg 5867  
Db 6309 TCCCTTTACCAAGAGACCTCTCGGTCGCGCCGCGAGCGGGCGCGGTGGAGCGTG 6368  
QY 5868 gccatgcttctctgcgaagccccagccccgcctcgcggtgagccccgcctctctcccca 5927  
Db 6369 GCCATGCTTTTACCAAAACCCCGGCGG-----GGAGCGCCCTCTCCACCCA 6419  
QY 5928 gcaagggctcgagcccccgccccctagtcctctgtctctgtggaacgccccactcgccc 5987  
Db 6420 GCAGAGTTCGAGCCCGATCTCTAGCACCCCGCCAGCTCCAGCCACACAGCCATCGGCC 6479  
QY 5988 gcaacctgcgaacactcgcaactcaccacgccccgagccccgcgcgcacactg 6047  
Db 6480 GCACCCAGCAAAACCTTGACCCCAACCTTGCCAGTCCGAGTCCGCGCGGCGGCGGCGG 6536  
QY 6048 cctcggtctcgagccccgcagcggaagactcaagtaaaccttttccatccagggaac 6107

Db 6537 CCTCGGCTCAGATCTGCACCGAGAAAAGACTCAAGATAAACCTTTTTCATCCAGGAAT 6596  
QY 6108 tggaaactcgttctctggtttaccac-----ggcagcagctacagccccgaaggggtgagc 6164  
Db 6597 TGGAACTCCGTCTCTGGGTTTACCACAGTGGAGCTGGGTACAGCCCCGATGGGTGGAGC 6656  
QY 6165 cgtcagccctgtgagctcaccagctgacccagcaaggggctcccccaagcactcg 6224  
Db 6657 CCATCAGCCCGGTGAGCTTCCCCAGCCTTGACCGACAAGGGGCTCTCCAAACCTCTGG 6716  
QY 6225 aagagctgacaaagcagccactggaggggagctgcccgaagcagcagcagccccgtga 6284  
Db 6717 AAGAGCTAGAGAAGAGCCACTTGGAAAGGGAGCTGCGCGCAACAAGCAGCAGGCCCATGA 6776  
QY 6285 agcttggcgggagggcgcccacactccacactcgcccgctgctgagagcagccct 6344  
Db 6777 AGCTCAGCGCGGAGGCTGCCATCTCCACATCTCTCGGCCACTTGCCTGAGAGCCAGCCT 6836  
QY 6345 cgtccagccctgctccagaccgccaggggtcaaggtcaccagcggtggtcacc 6404  
Db 6837 CATCAGGCCCACTCTCCAGACTGCCCCAGGCATCAAAAGGTCAACGAGGGTGTGTACCC 6896  
QY 6405 tggccagcacatcagtgaggtcatcacacaggactacacccggcaccacccacacagcgc 6464  
Db 6897 TGGCTCAGACATCAGCGAGGTCTATTACGACGACTACACGCGCCACCACCGCAGCAGC 6956  
QY 6465 tcagcgacccctgcccgcgcctctactctctctctctctctctctctctctctctctctg 6524  
Db 6957 TCAGTGGCCCTTCCCGCCCTCTCTACTCTTTCCTGGAGCCAGCTGCTCTGCTGG 7016  
QY 6525 acctccgcgcgcacccactgactctctctccgcgcgcgcgcgcgcgcgcgcgcgcgc 6584  
Db 7017 ATCTTCGCGGCCACCCAGTGAACCTTACCTTCCACCCCGCCGACCATGCGACCCAGCC 7076  
QY 6585 gtgctccccccacaggaagggggaaggtctccagagcccaaacagcgtcggtct 6644  
Db 7077 GGGATCCCCCACAGTGAAGGGGGCAAAAGTCCCCAGAACCCAGCAAAACATCGGTCC 7136  
QY 6645 tgggtggtgagagcggtattaaactgtctccacagcagggagcagcagcagcagcag 6704  
Db 7137 TGGCAGCAGCGAGGATGCCATGTAGCCTGTGTCCTCCACAGAGGGCATGTAGCCAG 7196  
QY 6705 ggcactcccgagtgctgtgtaccctgctgtaccggatggggaacagcagcgagccca 6764  
Db 7197 GACATGCTCGGAGCAGTGCCTACCCACTGCTGTATCGAGACGGGGAACAGGCGGAGCC- 7254  
QY 6765 gacgatgggtccaagtctccagggcaacacacagccagcccgcccttcttcagcaagc 6824  
Db 7255 -CAGGATGGGTCTAGAGTCTCCAGGCAACACCGCCAGCCAGCCCTTCTTCAGTAAGC 7313  
QY 6825 tgcagagagcaactcccgccatggtcaagtccaagaagaagagatcaacaagaagctga 6884  
Db 7314 TGACTGAGAGCACTCCGCCATGTGAAGTCGAGAAGCAGGAGATCAACAAGAATCA 7373  
QY 6885 acaccacaacgggaatgagcctgaatacata toagccagcctgggagcagatcttca 6944  
Db 7374 ACACCCACAACGGAAACAGGCGCAGAAATACAAATATTTGGCGAGCTGGGACGGAAATCTTCA 7433  
QY 6945 atatgcccgcctcacgggaacagcccttatgacctatagaagcagcggtgcaagaaac 7004  
Db 7434 ACATGCCCCCATCCTGAGGAGCGCTTATGACCTGTAGAAGCCAGCGGTGCAAGAAC 7493  
QY 7005 atgcagcaccacaatggggtgagggcctaataatagaagggcactcatgggttaaatg 7064  
Db 7494 ACGCCAGCACCAATGGGGCTAGAGGCCATTTATAGAAAGGCACTCATGGGTAAATG 7553  
QY 7065 accagtggaagagtgccccgcgcctcagcgccaatgtctttaacctctgaatgcagtg 7124  
Db 7554 ATCAGTGGAAAGAGCCCCCGGCTCGCGGCAATGCTTTTAAACCTCTGAATGCCAGCG 7613  
QY 7125 ccagcctgccc-----gctgctatgccataacgcgtgctgacgagcgagtgaccacac 7181















Db 4803 AGAGGCATTATCGACCTGTCCCAAGTGCACACCTGCGCGTGTGTGCCACCAACGCCA 4862  
QY 5246 ggcacccagcacccgcctgagacgccttgctactctccacacgcgcgcaccccttc 5305  
Db 4863 GGCACCCCTGCGACCGCATGACCGCTGCGCTACCTCCCACTGCGCCCGCCCACTTC 4922  
QY 5306 agcagccgccagcagctccccactctccccagagagtgccaaacacttgacaaaacca 5365  
Db 4923 AGCAGCCGCCACAGTAGCTACCGCTGTCCCCAGAGGCCCCACTCAGCTAGCTAAACCA 4982  
QY 5366 accacacgtctctcgcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 5425  
Db 4983 ACTGCACATCTTCACTCGAGCGG-----GAACGGGAACGTGAGCGG 5024  
QY 5426 ggcgggaaaaagtcacatccacacagcagctggagcagcagcagcagcagcagcagcagc 5485  
Db 5025 GAACGAGACAGTCCATCTCAGCTCTACCACTGAGCTGAGAGCATGACCCATCTGGAGA 5084  
QY 5486 cctgttacagagcagcagcgagcagcgagcagcgagcagcgagcgagcgagcgagcgagc 5545  
Db 5085 CTGTGTACGGAGCAGCAGCGGGGCTGGGGGACGACGCG----- 5125  
QY 5546 agcgcgcgcgcctccactccactccactccactccactccactccactccactccactcc 5605  
Db 5126 -----CCCGCGCTCCACACCCACCGACGACTCGCCCATCTCCCGCGGACCCGAG 5174  
QY 5606 gatccctccagcagacccagtgcttcacacacagcagcagcagcagcagcagcagcagc 5665  
Db 5175 GACGCTTGACAGCAGAGGCCAGTGTGTGCACAAACAGCAGATGAAGGGCGTGTACAC 5234  
QY 5666 gctgtggagccagcaagccacagctctg-----agttccacctccacctccctca 5716  
Db 5235 TCCGTGGAACCGCGCACCGCCAGCGTCTGAGTGGGCCAGGTCCACCTCCACCTCTTTCG 5294  
QY 5717 cccgttcgcaagctgcacattccactccactccactccactccactccactccactccact 5776  
Db 5295 CCGTGCCTCCAGCTGCGCACATTCACCTGCGCACCTGCGCACCTGCGCACCTGCGCACCT 5354  
QY 5777 gatggggtaccctaccctacagcagcagcagcagcagcagcagcagcagcagcagcagc 5836  
Db 5355 GAAGGGGTCTACCTTACCTCATGAGAGCGCGTCTCTGTACCAAGGAGACCTCTCGGGTC 5414  
QY 5837 gccgcgcagagcgcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5896  
Db 5415 GCGCGCGCGAGCGCGCGCGGTGGACGCTGGCCATGCTTTTACCAAAACCCCGCGGCG 5474  
QY 5897 cgtccgggctggagc 5956  
Db 5475 CG-----GGAGCGCGCGCTCTCACCACGCAAGAGCTCCGAGCGCGCGATCCCTAGCA 5525  
QY 5957 cctcctgtctgtgagcgcgcacacatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6016  
Db 5526 CCCCCAGCTCAGGCCACACAGCAGCTCGCGCGACCCCGACCCAGAAAGAACTTGGACCC 5585  
QY 6017 cagc 6076  
Db 5586 CATGCCAGTCCGGACCGCGCGCGG--CCACCTCGGGCTCAGATCTGCACCGAGAAAG 5642  
QY 6077 actcaagtaaacctttccatccaggaactggaactccgttctctggggttaaccac--- 6133  
Db 5643 ACTCAAGTAACCTTTTCTCATCAGGAATTGGAACCTCCCTTCTCTGGGTACACAGT 5702  
QY 6134 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6193  
Db 5703 GGAGCTGGCTACAGCCCCGATGGGGTGAGCCCATACAGCCGGTGAGCTCCCGGAGCGCTG 5762  
QY 6194 accacgacaaagggtcccaagcacttggaagcttcgacacagagcagcagcagcagcagc 6253  
Db 5763 ACCCAGCAAGAGGGCTCTCCAAACCTCTGGAAGAGCTAGAGAAGAGCCACTTGAAGGG 5822  
QY 6254 gagctgggcccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6313

Db 5823 GAGCTGGGCAAAAGCAGCCAGGCCAGGCCATGAAGCTACGCGGGAGGCTGCCATCTCCCA 5882  
QY 6314 cactggggcgctgcctgagagcagccctcgtccagccgcgtgctccagacgcgcgcgcgcgc 6373  
Db 5883 CATCTGGGGCAGCTGCGCGAGAGCCAGCCCTCATCCAGCCACCTCCTCCAGACTGCCCCA 5942  
QY 6374 ggggtcaaaagtccaccagcggtggtcacctcgtgccagacacacacagtgaggtacataca 6433  
Db 5943 GGCATCAAGGTCAACAGAGGTTGTTACCTTGCTCAGCACATCAGGAGGCTCATYACG 6002  
QY 6434 caggactacacccggcaccacccacagcagctcagcagcagcagcagcagcagcagcagcagc 6493  
Db 6003 CAGGACTACAGCGCCACCCAGCAGCTCAGTGGCCCTTCCGCGCCCTCTCTAC 6062  
QY 6494 tccctccctggggcgagctcccgctcgtgacacccgcgcgcgcgcgcgcgcgcgcgcgcgc 6553  
Db 6063 TCCCTTCCGGAGGACGCTGCCCTGTCTGTGATCTTCCGCCGCCACCCAGTGAGCTCTAC 6122  
QY 6554 ctccgc 6613  
Db 6123 CTCCACCCCGCGACCATGGCACCCAGCCCGGGGATCCCCCCACAGTGAAGGGGCAAA 6182  
QY 6614 aggtctccagagcacaacacagcagctcgtggtggtggtggtggtggtggtggtggtggtggt 6673  
Db 6183 AGGTCCCGCAGAACCCAGCAAAACATCGGTCTGGGCGACAGCGAGGATGCCATTTAGCT 6242  
QY 6674 gtgtcccccacgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6733  
Db 6243 GTGTCCCGCAGAGGATGACTGAGCCAGGACATGCTCGGAGACATGCTCGGAGACATGCT 6302  
QY 6734 ctgtaccggggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 6793  
Db 6303 CTGTATCGAGAGCGGGAACAGGGCGAGCG---CAGGATGGGTCTAGAGTCTCCAGGCAAC 6359  
QY 6794 accgagcagc 6853  
Db 6360 ACCAGCGCGCGCAACCTTCTCAGTAAGCTGACTGAGAGCAACTCGCGCATCGGTGAAG 6419  
QY 6854 tccagagcaagagatcaacaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 6913  
Db 6420 TCGAAGAGCAGAGAGATCAACAAACTCAACACCCCAACCCAGACCGAGCAGCAATAC 6479  
QY 6914 aatatcagcagc 6973  
Db 6480 AATATGGCCAGCGTGGAGCGAAATCTTCAACATGCCCGCATACACTGGAGCAGGCGCT 6539  
QY 6974 atgacctatagaacccagcggtgcaggaacatgcagcagcagcagcagcagcagcagcagc 7033  
Db 6540 ATGACCTGTAGAAGCCAGCGGTGCNAAGACACGCCACACCAACATGGGGCTAGAGGCC 6599  
QY 7034 ataattagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7093  
Db 6600 ATTATTAAGAGGCACATCATGGGTAAATATATGATCAGTGGGAAGAGCCCCCGCCCTCGGC 6659  
QY 7094 gccaatgcttttaacccctctgaatgccagtgccagcgcgcgcgcgcgcgcgcgcgcgcgcgc 7150  
Db 6660 GCCAATGCTTTTAAACCTCTGAATGCCAGCGCCAGTGTGCCCGCTGCTGTATGCCCAT 6719  
QY 7151 accgctgctgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7210  
Db 6720 ACCACTGCTCAGCAGCAGGAGTGACACGACCTACCTACCTCGCGAGGTGGAGGTGGAAAGCC 6779  
QY 7211 aagctcctggcagacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7270  
Db 6780 AAGGTCTGTGACAGACCTAGCAGCCGAAAGCCAAAGTGCAGCAGCAGGCTAGCGCTCC 6839  
QY 7271 ggggacggccacacccctctgtctcctcagtgacctcgagggagcagcagcagcagcagcagc 7330  
Db 6840 GGAGACCGACCCCTTCTGTCTCTCTAGTACACTCAGAGGGGAGTGCATTAATCGCCGAACA 6899  
QY 7331 ccgctcaccacacgc 7390  
Db 6900 CCATCCCAACCGTGTGTGGGAGGACCGGCCCTCATCTGCAGGGTCCAGGCCATTTCCC 6959



Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>



Qy 6346 gtccagcccgctgtccagaccgccccaggggtcaaaaggtcaccagcggtgtgtaacct 6405  
Db 612 gtccagcccgctgtccagaccgccccaggggtcaaaaggtcaccagcggtgtgtaacct 671  
Qy 6406 gcccagacatcagttaggttatacacagagactacacccggacccacacagagct 6465  
Db 672 gcccagacatcagttaggttatacacagagactacacccggacccacacagagct 731  
Qy 6466 cagcagccctgtcccgccccctctactctcctcctgggagcagctcccgctctgga 6525  
Db 732 cagcagccctgtcccgccccctctactctcctcctgggagcagctcccgctctgga 791  
Qy 6526 cctcgcgcgcacccagtgactctactcctccgccccggaccatggtgccccggcccg 6585  
Db 792 cctcgcgcgcacccagtgactctactcctccgccccggaccatggtgccccggcccg 851  
Qy 6586 tggctccccccacagcgaaggggcaagaggtctccagagccaaacaaagagtggttt 6645  
Db 852 tggctccccccacagcgaaggggcaagaggtctccagagccaaacaaagagtggttt 911  
Qy 6646 gggtaggtgagcaggtattgaacctgtaccggggtgggaacagagagcagccag 6705  
Db 912 gggtaggtgagcaggtattgaacctgtaccggggtgggaacagagagcagccag 971  
Qy 6706 gcactcccgagtgctgtaccgctgtgtaccggggtgggaacagagagcagccag 6765  
Db 972 gcactcccgagtgctgtaccgctgtgtaccggggtgggaacagagagcagccag 1031  
Qy 6766 caggatgggtcccaagtctccaggcaacacagcagccagccagctcttccagcaagct 6825  
Db 1032 caggatgggtcccaagtctccaggcaacacagcagccagccagctcttccagcaagct 1091  
Qy 6826 gaccagagcaactccgccaatgtcaagtccaagaaagaaagagatcaacaaagaaagctaa 6885  
Db 1092 gaccagagcaactccgccaatgtcaagtccaagaaagaaagagatcaacaaagaaagctaa 1151  
Qy 6886 caccacacacgggaatgagcctgaatacaataatcagcagcctgggagcgagatcttcaa 6945  
Db 1152 caccacacacgggaatgagcctgaatacaataatcagcagcctgggagcgagatcttcaa 1211  
Qy 6946 tatgcccgcatacccggaacaggccttatgacctatagaagccaggcgggtgcaggaaca 7005  
Db 1212 tatgcccgcatacccggaacaggccttatgacctatagaagccaggcgggtgcaggaaca 1271  
Qy 7006 tqcagacccaacatgggctggagggccataattagaaggaactcatgggttaatatga 7065  
Db 1272 tqcagacccaacatgggctggagggccataattagaaggaactcatgggttaatatga 1331  
Qy 7066 ccagtggaagagtgctcccgccgctcagcgccaatgcttttaacctctgaatgccagtg 7125  
Db 1332 ccagtggaagagtgctcccgccgctcagcgccaatgcttttaacctctgaatgccagtg 1391  
Qy 7126 cagcctgcccgtgctatgcccataaccgctgctgacgagcggagtgaccacacactcac 7185  
Db 1392 cagcctgcccgtgctatgcccataaccgctgctgacgagcggagtgaccacacactcac 1451  
Qy 7186 ctccgcaggtggcggaaggaagcgaaggtctctggaagccagcagccgaagaaagccaa 7245  
Db 1452 ctccgcaggtggcggaaggaagcgaaggtctctggaagccagcagccgaagaaagccaa 1511  
Qy 7246 gtccccggccccggcctggcatctgggagccggcgaacccctctgtctcctcagtgcactc 7305  
Db 1512 gtccccggccccggcctggcatctgggagccggcgaacccctctgtctcctcagtgcactc 1571  
Qy 7306 ggagggaactgcaaccccggaagccgctcacaacccgctgtggggagcagccctc 7365  
Db 1572 ggagggaactgcaaccccggaagccgctcacaacccgctgtggggagcagccctc 1631  
Qy 7366 gtccgaggttccagcctatccccataacccctgacatcgagctgagcgaggtgt 7425  
Db 1632 gtccgaggttccagcctatccccataacccctgacatcgagctgagcgaggtgt 1691

Qy 7426 catggtctcccaacccccacccggcctcccgcgggagcgccccctcgtcgtggccccca 7485  
Db 1692 catggtctcccaacccccacccggcctcccgcgggagcgccccctcgtcgtggccccca 1751  
Qy 7486 ccaagcctggagcagagagcccaagccaactcgtcgtcgcagtagcagagaactctcga 7545  
Db 1752 ccaagcctggagcagagagcccaagccaactcgtcgtcgcagtagcagagaactctcga 1811  
Qy 7546 cagcagtgactcagaacagcagcgggggcgggcgggcggtteuersperluatcaggtccc 7605  
Db 1812 cagcagtgactcagaacagcagcgggggcgggcgggcggtteuersperluatcaggtccc 1860  
Qy 7606 agcagccacaggaagccctcagggagcgggcggtcgcagactcccccaacaaagga 7665  
Db 1861 agcagccacaggaagccctcagggagcagggcggtcgcagactcccccaacaaagga 1920  
Qy 7666 aggaagccctcagtcgctcgcctccatcatctgtccgtcaagagccggcactcttg 7725  
Db 1921 aggaagccctcagtcgctcgcctccatcatctgtccgtcgcagagccggcactcttg 1980  
Qy 7726 cctgtctaaagccctaaactaaagactcccgccccggcctggccctgtcagaccttaactca 7785  
Db 1981 cctgtctaaagccctaaactaaagactcccgccccggcctggccctgtcagaccttaactca 2040  
Qy 7786 ggggatgtttacctgggtcgtcgggaagggagggcgggcgggcgggagggcgagcgag 7845  
Db 2041 ggggatgtttacctgggtcgtcgggaagggagggcgggcgggagggcgagcgagcgag 2100  
Qy 7846 gctgtgagcagcacaacagggcgggcgggcgggcgggcgggcgggagggcgagcgag 7905  
Db 2101 gctgtgagcagcacaacagggcgggcgggcgggcgggcgggcgggagggcgagcgag 2160  
Qy 7906 gcaactccagccactcctcccgccgaatgcatcttggaacaaagcttaaaactagctcg 7965  
Db 2161 gcaactccagccactcctcccgccgaatgcatcttggaacaaagcttaaaactagctcg 2220  
Qy 7966 cagcccccgccctccctcccgccctcccaatccgcttagcctctggagagatggagcga 8025  
Db 2221 cagcccccgccctccctcccgccctcccaatccgcttagcctctggagagatggagcga 2280  
Qy 8026 ggcctgtccagcccccaagtcgctcgttccgggtcccaacagactgccccagcgaacag 8085  
Db 2281 ggcctgtccagcccccaagtcgctcgttccgggtcccaacagactgccccagcgaacag 2340  
Qy 8086 atgctggaacccaagtcagccaggtggtggagcaaaagggcgaggtgccccctggggg 8145  
Db 2341 atgctggaacccaagtcagccaggtggtggagcaaaagggcgaggtgccccctggggg 2400  
Qy 8146 gaaaggtgctccagggactggactgttttttccacacatcgttgcgagcggtgggaa 8205  
Db 2401 gaaaggtgctccagggactggactgttttttccacacatcgttgcgagcggtgggaa 2460  
Qy 8206 ggaagggcagatgataatgagtggtgtttacaggggtatatttttatacactcaatga 8265  
Db 2461 ggaagggcagatgataatgagtggtgtttacaggggtatatttttatacactcaatga 2520  
Qy 8266 ataatcagatgttttacgaagaagagacttaccagtagtactcgtcgtgctttt 8325  
Db 2521 ataatcagatgttttacgaagaagagacttaccagtagtactcgtcgtgctttt 2580  
Qy 8326 gatctcgtctaccgttcaagagcgtgtgagcgccgagcagctggtgaccccaactcag 8385  
Db 2581 gatctcgtctaccgttcaagagcgtgtgagcgccgagcagctggtgaccccaactcag 2640  
Qy 8386 cagaccaaagggcgggagcgtgctg-tcaagccccgctgtgctcctcctccctccctt 8444  
Db 2641 cagaccaaagggcgggagcgtgctg-tcaagccccgctgtgctcctcctccctccctt 2700  
Qy 8445 cctgggcagagaatgaatcagatcgtattctgtggcgcccaatttgcagaggtgtggtga 8504  
Db 2701 cctgggcagagaatgaatcagatcgtattctgtggcgcccaatttgcagaggtgtggtga 2760  
Qy 8505 tctgtcatctacacagcgtgttctaataaaagcgaaattataactcccaaaaaaa 8564



```
|||||
Db 2761 TTCTGCTATTACACACGTCGTTCTTAATTAATAAAGCGAATTATATATCCAAAAA 2820
QY 8565 aaaaaa 8571
Db 2821 AAAAAA 2827

RESULT 8
S83390 2930 bp mRNA PRI 28-MAR-1997
LOCUS T3 receptor-associating cofactor-1 [human, fetal liver, mRNA, 2930 nt].
DEFINITION
ACCESSION S83390
VERSION S83390.1 GI:1911769
KEYWORDS human fetal liver.
SOURCE
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2930)
AUTHORS Sande,S. and Privalsky,M.L.
TITLE Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors
Mol. Endocrinol. 10 (7), 813-825 (1996)
JOURNAL 96408715
MEDLINE
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 179975] from the original journal article.
This sequence comes from Fig. 4.
FEATURES
source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
456..2765
/genes="T3 receptor-associating cofactor-1, TRAC-1"
456..2765
/genes="T3 receptor-associating cofactor-1, TRAC-1"
/notes="T3 receptor-associating cofactor-1, TRAC-1"
/notes="This sequence comes from Fig. 4; Authors report detecting several TRAC alternative transcripts; TRAC-1"
/codon_start=1
/product="T3 receptor-associating cofactor-1"
/protein_id="AAB50847.1"
/db_xref="GI:1911770"
/translation="MAQRADMLRGLSPRESSLALNYAAGPRGIDLSQVPHLPVLVPP
TPGTPATAMRLAYLPTAPQFSSRHSSSPLSPGPHLTKPTTSSSERERDRDRER
DRDRERKSILTTTVEHAPIWRPCTEQSSGSGSGSGSGSSRPASHSHAHQHP
ISRPTDALQORPSYLVNMGKLIITAVEPSTFVLKSTSTSPVRPAATFPATHGCP
LGTLDGVPVYTHMEPVLLPKPEARVARPERPRADTGHAFIAKPPARKSLPEAPSKG
SEPRPLVPVPSGHARTAPKALPHASPDPPAPASADPHREKTSKPFISQEL
ELRSLGVHSGSYSGEVPSPVSPSLTAPGVKHQVAVTIAOHISVITQDYTRHP
KLGEAAHLPLRLPESQSPSPSLQTAQVKGHRVVTIAOHISVITQDYTRHP
QQLSAPLPLYSFPGACSVLRLRPPSLDLYLPPDPHAPAGSPHSEGKRSPEPN
KTVSLGGEGDIEPVSPEGMTPEGHSRAVILLYRDGEQTEPSRHSKSPNTSQP
PAFFKLTNSAMVSKSKEINKLGNHNEPEYNISQPFIEFNPAITGTLMT
YRQAOVHEASTNNGLEAIIRKALMGKKGKAKVSFPRSKAKSPAPGLASGRPPSV
SVVHSEGDENRRPTLNRVWEDRPSASGTFPPYPLIMRLQAVMASPPPPGLPAGS
GPLAGAHAWDEEPKPLLCQSQYETLSDSE"
BASE COUNT 586 a 1149 c 780 g 415 t
ORIGIN

Query Match 30.6%; Score 2618.6; DB 97; Length 2930;
Best Local Similarity 94.6%; Pred. No. 0;
Matches 2786; Conservative 3; Mismatches 9; Indels 146; Gaps 2;

QY 4785 cccgcacagcacgctccgcagcaccacaccccccatctcgccctatgagcacctgc 4844
|||||
Db 133 CCCGCGACAGCACCGTCCCGGAGCACACACACACCCCATCTCGCCCTATGAGCACCTGC 192

QY 4845 ttcgggcgtagtgcgtagacctgtatgcagccacatccccctggccttcgacccca 4904
|||||
Db 193 TTCGGGGCGTAGTGGGTGAGACCTGTATGCGAGCCACATCCCCCTGGCCTTCGACCCCA 252
```



Qy	5985	ccgcaccccctgcgaagaacctgcactcaccacgccccagccccggagaccccgccggcgccac	6048
Db	1333	CCGCGACCCCCTGGAGAAACACTCGCACCTTCACACGCCAGCACCOCGCAGCCOCGC	1392
Qy	6045	ctgcctcgccctcgagccgcaccgcaggaaagactcaaatgtaaaccccttttcatccagg	6104
Db	1393	CTGCCTCGGGCCTCGGACCCGACCGGGAAAAGACTCAAAAGTAACCCCTTTTCCATCCAGG	1452
Qy	6105	aactggaaactccgttctcttggtttaccacggcagcagtacagccccgaagggtggagc	6164
Db	1453	AAC TGGNACTCCGTCTCTTG GTT ACC AGCG GAG CAG ACT ACAG CCC CGA AGGG GTG AGC	1512
Qy	6165	ccgtcagccctgtgagctcacccagctctgacccccagcaagggtcccccaagcacctgg	6224
Db	1513	CCGTACGCCCTGTGAGCTCACCCAGTCTGACCCACGACAAGGGCTCCCCAACGACCTGG	1572
Qy	6225	aagagctcgacaagagccaccttgaggggggagctcgccccaaagcagccagccccctga	6284
Db	1573	AAGAGCTCGAAGAAGACCACTTGGAGGGGGAGCTCGCGCCCAAGCAGCAGCCOCCGTGA	1632
Qy	6285	agcttggcgggagggccgccacctccacacctcgccgctgcccctgacctgaaagccagccct	6344
Db	1633	AGCTTGGCGGGAGCGCCGCCACCTCCACACTCGCGCCGCTGCCTGAGAGCCAGCCCT	1692
Qy	6345	cctcagccccgtgctccagacgccccagggtcaaaaggtcacagcgggtgggtcacccc	6404
Db	1693	CGTCCAGCCCGCTGTCCAGACGCGCCCAAGGGGTCAAAGGTCACCGGGGTGTCAACC	1752
Qy	6405	tggccagcacatcagtgagggtcatacacagagactacaacccggcaccaccaagcagc	6464
Db	1753	TGCGCCAGCACATCAGTGAGGTCTACACAGAGACTACACCGCGCACCCACACAGCAGC	1812
Qy	6465	tcagcgaccccttgcgcgcccccttacctcttcccttgggcgagctgccccgtcctgg	6524
Db	1813	TCAGCGCACCCCTGCGCCGCCCCCTCTACTCTCTCCCTTGGGGCCAGCTGCCCCGCTCTGG	1872
Qy	6525	acctcgcgcgccccaccagtgacctctacctccgcgccccggagccaattgtgccccggccc	6584
Db	1873	ACCTCGCGCGCCACCCAGTAGACTCTACTCTCGCGCCCGCGGACCATGTTGCCCCGCGCC	1932
Qy	6585	gtggctcccccaaagcgaaggggcgaagaggtctcagaagcctcaaaaagaagctcggtct	6644
Db	1933	GTGGCTCCCCCCACACGAAGGGGGCAAGAGGTCTCCAAGGCCAAACAAGACGTCGTCT	1992
Qy	6645	tgggtggttgagagcgtattgaacctgtgtccccaccggagggatcacagagccag	6704
Db	1993	TGGGTGGTGGTGAAGCGGTATTGAACCTGTGTCCCCACCAGGAGGGCATGACGGAGCCAG	2052
Qy	6705	ggcactccoggagtgctgtgaccgcgtgctgtagccggatgagggaacagacgagccca	6764
Db	2053	GGCAC'TCCCGAGTGTGTGT'ACC'CGCTGCTGT'ACC'GGAT'GGGGAACAGACGAGGCCA	2112
Qy	6765	gcaggatgggtccaaagtctcagcgcaacacacagccagccagccagccttcttaagaagc	6824
Db	2113	GCAGAT'GGGCT'C'AAGTCT'CCAGC'CAACACAGCCAGCCGCGCCAGCCTTCTTCAGCAAGC	2172
Qy	6825	tgaccagagcaactccgcctatgttcaagtccaagaagcaagagatatcaaaaaagctga	6884
Db	2173	TGACCGNAGCAACTCCGCCATGGTCAAGTCCCAAGCAACAGAGATCAACAAGAAGCTGA	2232
Qy	6885	acaccacaacgggaatgagcctgaatacaatatcagccagcctgggacggagatcttca	6944
Db	2233	ACACCCACAACCGGAATGAGCTGAATACAATATCAGCCAGCCCTGGGACGGAGATCTTCA	2292
Qy	6945	atatgcccgccatcacccggaacagcccttatgacctatagaagccagcgcggtgcaggaac	7004
Db	2293	ATATGCCCGCCATCACCGGAACAGCGCCTTATGACCTATATAAGACCGACGCGGTGCAGAAC	2352
Qy	7005	atgccagcaccacaatgggcttggaggccataattagaagagcactcatgggtaaatatg	7064
Db	2353	ATGCCAGCACCAATATGGGGCTTGGAGGCCATANTTAGAAAGGCACTCATG-----	2402
Qy	7065	accagtgggaagagtcctccgcgcgtcagcgccaatgctttttaacccctctgaatgccagt	7124

[illegible]

RESULT	9
AC073916/c	
LOCUS	AC073916 209594 bp DNA 12-MAR-2001
DEFINITION	Homo sapiens chromosome 12q clone RP11-408I18, WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION	AC073916
VERSION	AC073916.23 GI:13270491
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 209594)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaraturunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,



Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, K., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kreshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisleged, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 209594)  
 Worley, K. C.

Submitted (05-JUL-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Mar 10, 2001 this sequence version replaced gi:12965246.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HBW  
 Center clone name: RP11-408118  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-primer Bodipy: 24% of reads  
 Chemistry: Dye-terminator Big Dye: 76% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 209115 bases at least Q40  
 Consensus quality: 213878 bases at least Q30  
 Consensus quality: 216310 bases at least Q20  
 Estimated insert size: 205541; sum-of-ontigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 6.2x in Q20 bases; sum-of-ontigs estimation  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: this is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 49656: contig of 49656 bp in length  
 \* 49657 49756: gap of unknown length  
 \* 49757 104158: contig of 54402 bp in length  
 \* 104159 104258: gap of unknown length  
 \* 104259 148912: contig of 44654 bp in length  
 \* 148913 149012: gap of unknown length  
 \* 149013 172642: contig of 23630 bp in length

\* 172643 172742: gap of unknown length  
 \* 172743 200085: contig of 27343 bp in length  
 \* 200086 200185: gap of unknown length  
 \* 200186 205657: contig of 5472 bp in length  
 \* 205658 205757: gap of unknown length  
 \* 205758 206934: contig of 1177 bp in length  
 \* 206935 207034: gap of unknown length  
 \* 207035 208394: contig of 1360 bp in length  
 \* 208395 208494: gap of unknown length  
 \* 208495 209594: contig of 1100 bp in length.

FEATURES  
 Location/Qualifiers  
 1..209594  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12q"  
 /clone="RP11-408118"  
 BASE COUNT 42687 a 58364 c 60674 g 47056 t 813 others  
 ORIGIN

Query Match 13.2%: Score 1132.4; DB 75; Length 209594;  
 Best Local Similarity 97.7%: Pred. No. 1.8e-159;  
 Matches 1168; Conservative 3; Mismatches 11; Indels 13; Gaps 2;

QY 7368 ccgcaggttccacgcccattccctcaacacccctgatcatcgctgcaggcggtgtca 7427  
 |||||  
 Db 90599 CGCAGGTTCCACGCCATTCCCTACAAACCCCTGATCATGGCTGCAGGGGGGTCA 90540  
 |||||  
 QY 7428 tggcttcccccacccacccgggctcccccggggcaggcggccctcgtgccccacc 7487  
 |||||  
 Db 90539 TGGCTTCCCCACCCACCGGGCTCCCGGGGAGGGGGGGCTCGCTGCCGCCACC 90480  
 |||||  
 QY 7488 acgcttggagcagagcccaagccactgctctcgcagtcagacacactctccgaca 7547  
 |||||  
 Db 90479 AGCCCTGGAGCAGAGAGCCCAAGCCACTGCTCTGCTGCGAGTACGAGACACTCTCGACA 90420  
 |||||  
 QY 7548 gcgagtgcactcagaacagggcgggggggggcgggcggtggttcagggtccag 7607  
 |||||  
 Db 90419 GCGAGTGACTCAGAACAGGGGGGGGGGGG-----GGCGGTGTCAAGGTCCCAG 90372  
 |||||  
 QY 7608 cgagccacaggaacggccctgcaggagcggggcggtcgtccgactcccccaacagggaag 7667  
 |||||  
 Db 90371 CGAGCCACAGGAACGGCCCTGCAGGAGAGGGCGGCTGCCGACTCCCCCAACAAAGGAAG 90312  
 |||||  
 QY 7668 gagcccttcagtcgcctcgcctccatctcctcgcagtcagcgggacatctgcc 7727  
 |||||  
 Db 90311 GAGCCCTTCAGTCCGCCCTGCGGCTCCATCTCTGCTGCTGCGAGCGCGGATCTTGCC 90252  
 |||||  
 QY 7728 tgcctaaagccttaactaaagactcccccggggcggtggtgaccttactcagg 7787  
 |||||  
 Db 90251 TGTCTAAAGCCTTAACCTAAGACTCCCGCCCGGGCTGGCCCTGTGCAGACCTTACTCAGG 90192  
 |||||  
 QY 7788 ggaattaccttgcttcgggaaggggggaaaggggggggggggggggggggggggggggg 7847  
 |||||  
 Db 90191 GGATCTTTACCTTGGTGTCTCGGGAAGGGGGAAGGGGGGGGGGGGGGGGGGGGGGGGGGG 90132  
 |||||  
 QY 7848 gtgtgagccacacacagggcggccagggcggccagggcggccagggcggccagggcggccaggg 7907  
 |||||  
 Db 90131 GGTGGCAGCCACACGCGAGGG 90072  
 |||||  
 QY 7908 acctcagccactgcctccccccgaatgcatttgaaacaaagtctaaactgagctcgca 7967  
 |||||  
 Db 90071 ACCTCCACGCCACTGCCTCCCGGAATGCAITTTGAACCAAGAGTCTAACTGAGCTCGCA 90012  
 |||||  
 QY 7968 gccccggcgcctccctccgctcccatcccgcttagcgctcttgacagatggagcaggg 8027  
 |||||  
 Db 90011 GCCCCGGCGCCTCCTCCCGCTCCCATCCCGCTTAGCGCTCTGGACAGATGGAGCGCAGG 89952  
 |||||  
 QY 8028 ccctgtccagccccagtcgcttccgttccccacagactccccagcccaacagat 8087  
 |||||  
 Db 89951 CCCTGTCCAGCCCCCAGTGCCTCTGTTCCGGTCCCGACAGACTGCCCGCCCAACAGAGAT 89892  
 |||||  
 QY 8088 tgctggaacacaaagtcaggccaggtggggcggaacaaagggccaggtgaggcctgggggga 8147







```
* 35215 35314: gap of 100 bp
* 35315 37181: contig of 1867 bp in length
* 37182 37281: gap of 100 bp
* 37282 39642: contig of 2361 bp in length
* 39643 39742: gap of 100 bp
* 39743 41978: contig of 2236 bp in length
* 41979 42078: gap of 100 bp
* 42079 44963: contig of 2885 bp in length
* 44964 45063: gap of 100 bp
* 45064 47265: contig of 2202 bp in length
* 47266 47365: gap of 100 bp
* 47366 49847: contig of 2482 bp in length
* 49848 49947: gap of 100 bp
* 49948 52474: contig of 2527 bp in length
* 52475 52574: gap of 100 bp
* 52575 54767: contig of 2193 bp in length
* 54768 54867: gap of 100 bp
* 54868 57938: contig of 3071 bp in length
* 57939 58038: gap of 100 bp
* 58039 60538: contig of 2500 bp in length
* 60539 60638: gap of 100 bp
* 60639 63422: contig of 2784 bp in length
* 63423 63522: gap of 100 bp
* 63523 66038: contig of 2516 bp in length
* 66039 66138: gap of 100 bp
* 66139 69185: contig of 3047 bp in length
* 69186 69285: gap of 100 bp
* 69286 72321: contig of 3036 bp in length
* 72322 72421: gap of 100 bp
* 72422 75240: contig of 2819 bp in length
* 75241 75340: gap of 100 bp
* 75341 77598: contig of 2258 bp in length
* 77599 77698: gap of 100 bp
* 77699 82485: contig of 4787 bp in length
* 82486 82585: gap of 100 bp
* 82586 86177: contig of 3592 bp in length
* 86178 86277: gap of 100 bp
* 86278 91129: contig of 4852 bp in length
* 91130 91229: gap of 100 bp
* 91230 94475: contig of 3246 bp in length
* 94476 94575: gap of 100 bp
* 94576 99382: contig of 3807 bp in length
* 99383 99482: gap of 100 bp
* 99483 103891: contig of 5409 bp in length
* 103892 103991: gap of 100 bp
* 103992 109749: contig of 5758 bp in length
* 109750 109849: gap of 100 bp
* 109850 116075: contig of 6226 bp in length
* 116076 116175: gap of 100 bp
* 116176 122437: contig of 6262 bp in length
* 122438 122537: gap of 100 bp
* 122538 132392: contig of 9855 bp in length
* 132393 132492: gap of 100 bp
* 132493 141881: contig of 9389 bp in length
* 141882 141981: gap of 100 bp
* 141982 161970: contig of 19989 bp in length.
```

## FEATURES

## source

```
1. 161970
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /map="12"
   /clone="RP11-665C13"
   /clone_lib="RPC1-11 Human Male BAC"
   1. 1052
   /note="assembly_fragment"
   1153. 2378
   /note="assembly_fragment"
   2479. 3812
   /note="assembly_fragment"
   3913. 5368
   /note="assembly_fragment"
   5469. 6519
```

```
misc_feature
6620..7852
/note="assembly_fragment"
misc_feature
7953..8464
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
8565..10347
/note="assembly_fragment"
10448..12145
/note="assembly_fragment"
12246..14006
/note="assembly_fragment"
14107..15358
/note="assembly_fragment"
```

Query Match 11.1%; Score 948.4; DB 71; Length 161970;  
Best Local Similarity 97.6%; Pred. No. 3.8e-132;  
Matches 965; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

```
Qy 7574 gggcgggggtteuersperlugtcaggtcccccagcagccacaggaacgagcctgcagga 7633
Db 60639 GCGGGGGGGGGGGGGGGTGTCTCAGGTCCAGGCGCCACAGGACGCGCCCTCGCAGGA 60698
Qy 7634 ggggggggctgcgactcccccaacgaaggagggcctgagtcgcgctgcgctcc 7693
Db 60699 GCAGGGCGGTTGCCGACTCCGCCCAACCAAGGAAGAGAGCCCTGAGTCCGCCCTGCCCTCC 60758
Qy 7694 atccatctgcgtccagagccggcatcttcctgctctaaagccttaactaagactccc 7753
Db 60759 ATCCATCTGTCGTCCAGAGCGGCATCCTTGCTCTGTCTAAAGCCTTAACTAAGACTCCC 60818
Qy 7754 gcccgggctgcctgtgcaaccttactcaggaaggtttacctgtgttcgagaagg 7813
Db 60819 GCCCCGGGTGCCCCCTGTGCAGACCTTACTCAGGGGATGTTTTACTGTGTGTCGGGAAG 60878
Qy 7814 gagggaagggggggggagggggcagcgctgtgtggcagccacacacagggcgagca 7873
Db 60879 GAGGGGAAGGGGCGGGGAGGGGCGCAGCGAGCGCTGTGGCAGCCACACAGCGCGGCA 60938
Qy 7874 gggcgccagggagcccaagcaggatgaccacgacccctccacgacactgcctccccccgaa 7933
Db 60939 GGGCGGCCAGGAGACCCAAAGCAGATGACCCAGCAGCTCCACGCCACTGCTCTCCCGGAA 60998
Qy 7934 tgcatttgaaacaaagtctaaactgagtcgcagcccccgccctccctccctccc 7993
Db 60999 TGCATTTGGAACCAAGTCTAAACTGAGTCTGCACGCCCGCGCCCTCTCTCCGCTCC 61058
Qy 7994 atcccgcttagcgtctggacagatggacgagggcctgtccagccccccagtcgctcgt 8053
Db 61059 ATCCCGCTTAGCGCTCTGGACAGATGGACGAGGCGCTGTCCAGCGCCCGCAGTGCCTGT 61118
Qy 8054 tccggtccccacagactccccccagcacaagatgtgtggaaaacaaagtcagggcaggtg 8113
Db 61119 TCCGCTCCCGCCAGACTGCCCGCAGCAACGAGATTGCTGTGAACCAAGTCAAGCCAGGTG 61178
Qy 8114 ggcggacaaaagggcgaggtgcgctggggggaacagatgctccgagagactggactgt 8173
Db 61179 GCGGGAACAAAGGGCGCAGGTGGGGCTGTGGGGGAACGAGATGCTCCAGGACTGGAGTGT 61238
Qy 8174 tttttcacacatcgttgcgcagcgggtgggaaggaaaagcagatgtaaatgatgtgttg 8233
Db 61239 TTTTTCACATCTGTTGCCGAGCGGTGGGAAGGAAGGAGATGTAATGATGTGTTGG 61298
Qy 8234 ttacaggggtattttttgatctaccttcaatgaattaatcagatgttttacgcaaggag 8293
Db 61299 TTTACAGGGTATATTTTGTATACCTTCAATGAATTAATTAATTTTACCAAGAGGAG 61358
Qy 8294 gacttaccagattactgctgctgtgttttgcctctgctctgcttaccgttcaagagcggtg 8353
Db 61359 GACTTACCAGATTATACGCTGCTGTGCTTTTGTGATCTCTGCTTACCGTTCAAGAGCGGTG 61418
Qy 8354 tgcaggcgacagtcgtgtgaccccatcactgcaggaacaaaggggggaggggactgctcg- 8412
```







Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L7063

Center clone name: 339\_B\_19

-----  
\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 964: contig of 964 bp in length  
\* 965 1064: gap of 100 bp  
\* 1065 2044: contig of 980 bp in length  
\* 2045 2144: gap of 100 bp  
\* 2145 3096: contig of 952 bp in length  
\* 3097 3196: gap of 100 bp  
\* 3197 4131: contig of 935 bp in length  
\* 4132 4231: gap of 100 bp  
\* 4232 5218: contig of 987 bp in length  
\* 5219 5318: gap of 100 bp  
\* 5319 6322: contig of 1004 bp in length  
\* 6323 6422: gap of 100 bp  
\* 6423 7388: contig of 966 bp in length  
\* 7389 7488: gap of 100 bp  
\* 7489 8466: contig of 978 bp in length  
\* 8467 8566: gap of 100 bp  
\* 8567 9527: contig of 961 bp in length  
\* 9528 9627: gap of 100 bp  
\* 9628 10612: contig of 985 bp in length  
\* 10613 10712: gap of 100 bp  
\* 10713 11685: contig of 973 bp in length  
\* 11686 11785: gap of 100 bp  
\* 11786 12725: contig of 940 bp in length  
\* 12726 12825: gap of 100 bp  
\* 12826 13794: contig of 969 bp in length.  
\* 13795 13894: gap of 100 bp  
\* 13895 14875: contig of 981 bp in length  
\* 14876 14975: gap of 100 bp  
\* 14976 15963: contig of 988 bp in length  
\* 15964 16063: gap of 100 bp  
\* 16064 17033: contig of 970 bp in length  
\* 17034 17133: gap of 100 bp  
\* 17134 18122: contig of 989 bp in length  
\* 18123 18222: gap of 100 bp  
\* 18223 19199: contig of 977 bp in length  
\* 19200 19299: gap of 100 bp  
\* 19300 20275: contig of 976 bp in length  
\* 20276 20375: gap of 100 bp  
\* 20376 21340: contig of 965 bp in length  
\* 21341 21440: gap of 100 bp  
\* 21441 22409: contig of 969 bp in length  
\* 22410 22509: gap of 100 bp  
\* 22510 23478: contig of 969 bp in length  
\* 23479 23578: gap of 100 bp  
\* 23579 24528: contig of 950 bp in length

24529 24628: gap of 100 bp  
24629 25614: contig of 986 bp in length  
25615 25714: gap of 100 bp  
25715 26685: contig of 971 bp in length  
26686 26785: gap of 100 bp  
26786 27766: contig of 981 bp in length  
27767 27866: gap of 100 bp  
27867 28836: contig of 970 bp in length  
28837 28936: gap of 100 bp  
28937 29940: contig of 1004 bp in length  
29941 30040: gap of 100 bp  
30041 31073: contig of 1033 bp in length  
31074 31173: gap of 100 bp  
31174 32163: contig of 990 bp in length  
32164 32263: gap of 100 bp  
32264 33255: contig of 992 bp in length  
33256 33355: gap of 100 bp  
33356 34343: contig of 988 bp in length  
34344 34443: gap of 100 bp  
34444 35421: contig of 978 bp in length  
35422 35521: gap of 100 bp  
35522 36507: contig of 986 bp in length  
36508 36607: gap of 100 bp  
36608 37558: contig of 951 bp in length  
37559 37658: gap of 100 bp  
37659 38645: contig of 987 bp in length  
38646 38745: gap of 100 bp  
38746 39754: contig of 1009 bp in length  
39755 39854: gap of 100 bp  
39855 40828: contig of 974 bp in length  
40829 40928: gap of 100 bp  
40929 41925: contig of 997 bp in length  
41926 42025: gap of 100 bp  
42026 43035: contig of 1010 bp in length  
43036 43135: gap of 100 bp  
43136 44097: contig of 962 bp in length  
44098 44197: gap of 100 bp  
44198 45179: contig of 982 bp in length  
45180 45279: gap of 100 bp  
45280 46245: contig of 966 bp in length  
46246 46345: gap of 100 bp  
46346 47341: contig of 996 bp in length  
47342 47441: gap of 100 bp  
47442 48397: contig of 956 bp in length  
48398 48497: gap of 100 bp  
48498 49441: contig of 944 bp in length  
49442 49541: gap of 100 bp  
49542 50510: contig of 969 bp in length  
50511 50610: gap of 100 bp  
50611 51575: contig of 965 bp in length  
51576 51675: gap of 100 bp  
51676 52628: contig of 953 bp in length  
52629 52728: gap of 100 bp  
52729 53677: contig of 949 bp in length  
53678 53777: gap of 100 bp  
53778 54756: contig of 979 bp in length  
54757 54856: gap of 100 bp  
54857 55849: contig of 993 bp in length  
55850 55949: gap of 100 bp  
55950 56913: contig of 964 bp in length  
56914 57013: gap of 100 bp  
57014 57990: contig of 977 bp in length  
57991 58090: gap of 100 bp  
58091 59058: contig of 968 bp in length  
59059 59158: gap of 100 bp  
59159 60147: contig of 989 bp in length  
60148 60247: gap of 100 bp  
60248 61223: contig of 976 bp in length  
61224 61323: gap of 100 bp  
61324 62301: contig of 978 bp in length  
62302 62401: gap of 100 bp  
62402 63354: contig of 953 bp in length  
63355 63454: gap of 100 bp



```
* 63455 64436: contig of 982 bp in length
* 64437 64536: gap of 100 bp
* 64537 6519: contig of 983 bp in length
* 6520 65619: gap of 100 bp
* 65620 65537: contig of 918 bp in length
* 65538 66637: gap of 100 bp
* 66638 67612: contig of 975 bp in length
* 67613 67712: gap of 100 bp
* 67713 68661: contig of 949 bp in length
* 68662 68761: gap of 100 bp
* 68762 69741: contig of 980 bp in length
* 69742 69841: gap of 100 bp
* 69842 70831: contig of 990 bp in length
* 70832 70931: gap of 100 bp
* 70932 71890: contig of 959 bp in length
* 71891 71990: gap of 100 bp
* 71991 72991: contig of 1001 bp in length
* 72992 73091: gap of 100 bp
* 73092 74042: contig of 951 bp in length
* 74043 74142: gap of 100 bp
* 74143 75119: contig of 977 bp in length

Query Match      8.5%; Score 732.8; DB 73; Length 79494;
Best Local Similarity 93.3%; Pred. No. 4.8e-100;
Matches 839; Conservative 3; Mismatches 49; Indels 8; Gaps 8;

Qy 7576 ggcggcggtuersperluatcaggtccagcagcacagagcagccctgcaggagc 7635
Db 5175 GcCCGGGGGGGGCGGTTTCAGGTCCCGAGGAGCACCA-GAACGGCCCTTCAGAAAGC 5117
Qy 7636 gggcggtgcgcactcccccaacaaagagagccctgagtcgcctgcgcctccat 7695
Db 5116 AGGCGGTTCCNTCTCCCCACACACCCAGGA-GAGCCCTTGAGTCCGCTGGCNCAT 5058
Qy 7696 coactgtccgtccagacc-ggaactctgctctctaaagccttaactaagactccg 7754
Db 5057 CCATNTGTCCTCCAGAGCGGGGATCTTTCGCTGTTACAGCCTTANTAGACCTCCG 4998
Qy 7755 cccgggtgcctctgcaacacttactcagggatgtttacatgtgtgctcggaagg 7814
Db 4997 CCCCCGGTGGCCCTGTGCAGACCTTACTCAGNGGATG-TTACCTGGTGCCCG-GCAGNG 4940
Qy 7815 aggggaagggcgggggggcgagcgagcggtgtgtgagcagcacacacagcgggccag 7874
Db 4939 AGGGAAGGGCGGGAGGGGCGACGCGGTGCGAGCCACACGACGACGCGCCAG 4880
Qy 7875 ggcgcagggagaccacaaagcagggatgacacagcaactccagcgaactgcctcccccgaat 7934
Db 4879 GCGGCCACAGGACCCAAAGCAGGATGACACGACCTCCACGCCACTCCCTCCCCCGAAT 4820
Qy 7935 gatttgaacaaagtctaaactgagctgcagcccccgccctccctccgctccca 7994
Db 4819 GCATTTGAACAAAGTTAAATGAGTTCGAGCCNCCGCGCCCTCCCTCCGCTCCCA 4760
Qy 7995 tccgcttagcgtctgacagatgagcagcgccctgtccagcccccaagtcgctcgtt 8054
Db 4759 TCCCGCTTAGCGCTCTGACAGATGAGCAGCGCCCTGTCCAGCCCGCCAGTCGCTCGT 4700
Qy 8055 ccggtcccccagactgcccagcacaagatgtctgagaaacaaagtcaggccaggttg 8114
Db 4699 CCGGTCACAGATGCCCCAGCCAAAGCAGATTGCTGGAAACCAAGTCAGGCCAGGTGG 4640
Qy 8115 gcggacaaagggcgagtgcgccctgggggaaacggagtgctccgaggaactggtt 8174
Db 4639 GCGGACAAAGGGCCAGGTGCGGCTTGGGGGAACGGATGCTCCGAGACTGGAGTGT 4580
Qy 8175 tttcacacatogtttgcgcagcggtgggaagaaagcagatgtaaatgtgttgg 8234
Db 4579 TTTTTCACATCGTTGCCGACGCGGTGGGAAGGAAGCAGATGTAATGATGTGTGGT 4520
Qy 8235 ttacaggtatattttgatacccttcaatgaataataatcagatgtttcaccaaggaag 8294
Db 4519 TTACAGGTATATTTTGTATACCTTCAATGAATTAATTCAGATGTTTTTACCAAGGAAG 4460
```

```
Qy 8295 attaccagcttactgctgtgtgttttgatctctgcttaccgttcaagagcggtgt 8354
Db 4459 ACTTACCAGTATTACTGCTGCTGTCTTTGATCTGCTTACCGTTCAGAGCGGTGT 4400
Qy 8355 gcagccgacagctggtgagcccatcactgcagacaaagggcggggactctgct -t 8413
Db 4399 GCAGGCCGACAGTCGGTGCACCCCATCTCGCAGGACCAAGGGGGGGGACTGCTGGCT 4340
Qy 8414 cacgcccgcgtgtctcctccctccctccctccctccctccctccctccctccctccat 8472
Db 4339 CACGCCCGGCTGTGCTCCCTCCCT -CCTTCTTGGGCAGAA-GAATTCGATCGTAT 4283

RESULT 13
AF044209 Homo sapiens nuclear receptor co-repressor N-CoR mRNA, complete cds.
LOCUS AF044209 7940 bp mRNA PRI 01-SEP-1998
ACCESSION AF044209
VERSION AF044209.1 GI:3510602
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 7940)
Wang,J., Hoshino,T., Redner,R.L., Kajigaya,S. and Liu,J.M.
ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
transcription by interaction with the human N-CoR/msin3/HDAC1
complex
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (18), 10860-10865 (1998)
MEDLINE 98393736
REFERENCE
2 (bases 1 to 7940)
Wang,J. and Liu,J.M.
Direct Submission
JOURNAL Submitted (22-JAN-1998) Hematology Branch, National Heart, Lung and
Blood Institute, Bldg. 10, Rm. 7C103, 9000 Rockville Pike,
Bethesda, MD 20892, USA
FEATURES
source location/Qualifiers
1. .7940
/organism="Homo sapiens"
/db_xref="taxon:9606"
241..7563
/codon_start=1
/product="nuclear receptor co-repressor N-CoR"
/protein_id="AAC33550.1"
/translation="MSSSGYPPNQGAFSTEQSRYPHVSQYVTPNTRHQEFAVPDYR
SSLEVSQAQLQQQQQQQLRRRPLLSSEFHPGSDRQPERRTSYEPFHPGSPVDHD
SLSKRPLRQVSDSHFORVSAVPLVHPLEGLRASADAKKDPAGGKHEAPSSPI
SGQPCDDONASPSKLSKEELIOSMDRVDEIAKVEQOILKLLKKQOOLKEEAKPE
PEKVPSPPEVKHRSIVQIILYDENRKAEEAHKIFEGLPKVELPLYNQPSDTKYVH
ENIKNOVMKKLILFLRNHARKQREKICQYDQLMKAEWKVDRIENPNPKAK
ESTREYKQFPEIRKQRFQVQGGAGLSATIAEISEHSEIIDGLESEQN
NEQMROLSYIPPMFDAQRVKFINMGIMEDPMKYVDROFMNVTDRKEIFKD
KFTQHPNGLIASYLERSKVPDCLVLYLTNNKYNKALVRNRYKRRNQOILAP
SQEKEEKEEDKAEKKEKEKEDEKEDKENTKEKDIDGTAEEETEERQ
ATPRGKTANSQGRKGRITRSMTNEAASASAAAATEEPPPLPPPPPISTPEVE
TSWTETEEMVARKGLVHGHNAAATAMVGTSEAGCKNFYKRNKRLNDNLQOH
KQTSRRPREEDVSCQSVASTAQEDIEDASNEENPEDESVAVPQSDPEN
ATSRGNTPEAVELETETAPTSPSLAVETPAEDSVETQVNDISAEATQMDV
DOQHEAEEGSCVDPPTKADSDVDVVRPNHASKVEDNTKERDLRSKVEPR
DEDLVVAQINAPRPPQSDSSATCDAEDVDGPERQRMFPMDSKPLLNPTEGSI
LVSPKLPNLDLPOLHRAAVIPMWVCTPCNIPITPVSGYALYORHLKAMHESAL
LEOROROEIDLECSSTSCPTSKSPNREVLQAPHQLITNLPGLPTTRPT
RPPPLPSSTKTIVASEKPSFIMGSSIQCTPGTILTSHNQASITGTPSPVGSISL
GLPQESAKSATLPYIKQEEFSPRSQSQEGGLVRAQHEGVVGTAGTAIGSITR
GTPTSKISVESIPSLGSIQTQTPALPQTGIPTEALVKGSI SRMP IEDSPKGRER
ASKGHVIEKSGHILSYDNIKNAREGTRSPRTHAHSIKRSYVESVGNIKQGMRE
SPYSAPLEGILCRALPGSPHDLKERTVLGSSIMOGTPRATTFESPDGLKYKQIKR
ESPIRAFEGAITKRPYDGTWKEMGRSIIHEIPRODILTQBSRKTPEVQSTPRLI
EGSIPQCTPIKFDNNSQSKAIKNVKSILITGTPSKLSKRGMPLEIVPENIKVVERGKYE
```



Query Match	7.4%	Score 630;	DB 88;	Length 7940;
Best Local Similarity	65.1%	Pred. No. 2e-84;		
Matches 1015;	Conservative 0;	Mismatches 495;	Indels 48;	Gaps 4;
QY 491	cgccacgctgtccaagaggagctatccagaacatgcacgcgtggaccagagatc	550		
DB 757	CTTTCAAACCTCTCAAAGGAAGAGTTAATCAGAGTATGGATCTGTAGATCGAGAATT	816		
QY 551	accatggtagcagcagatctctaagctgaagaagaagcagcaacagctgggagagag	610		
DB 817	GCAAAAGTAGAACAGCAGATCCTTAAACTGAAAAGAAACAACAACAGCTTGAAGAAGAG	876		
QY 611	gctgccaagcgcgcgagcctgagaagccgctgtcaaccgcgcgcctatcgagtcgaagcac	670		
DB 877	GCAGCTAAACCTCTGTAGCCTGGAAGCCGTGTCCCTCTCTCTGTGTGAGCAGAAACAC	936		
QY 671	cgagcctggtgcagatcatctacgacgagaacccggaagagctgaagctgcacatcgg	730		
DB 937	CGCAGTATTGTCCNAATTATTTATGATGAGAAATCGGAAAAAAGCAGAAGAGCTCATAAA	996		
QY 731	attctggaagcctcgtggccccagtggtgagctgcgctgtacaaaccagcctccgcagacc	790		
DB 997	ATTTTGAAGGTCTTGCCCCCAAGATTTGAATGCCACTGTATTAACCGACCATCAGATACC	1056		
QY 791	cggcagtatcatgagaacatcaaaaataaaccaggcgtgcggaagaagactaatcttgtac	850		
DB 1057	AAGGTGTACCATTAGAAACATCAAGACAAACACAGGTGATGAGGAAAAAACTCATTTTATT	1116		
QY 851	tcaagaggaggaatcacgctcggaacaatggaagcagagcttctccagcgctatgac	910		
DB 1117	TTTAAAGAGAAATCATGTGCAAGAAACAAGGGAACAATAATCTGCCGCGTTATGAT	1176		
QY 911	cagctatgaggccttggaaaaaagtggtgagcgcatcgaacaacccgcgcgcggcg	970		
DB 1177	CAGCTCATGGAGGCATGGGAGAAAAAAGTGACAGAAATGAAAAATAATCTCTCGCAGGAAA	1236		
QY 971	gccaagagagcaaggtgcgcgagactactcgaagaagcagttccctcgagatccgcaagcag	1030		
DB 1237	GCTAAGAAAGCAAAACAAGGGAATACTATTGAAAGCAGTTTCCAGAAATTCGAAAAACA	1296		
QY 1031	cgcgagctgcaggagcgcagagcagagtggtgggcagcgggcagtgggcttccatg	1090		
DB 1297	AGAGACAGCAAGAAAGATTTTCAG--CGAGTTGGGCAGAGGGAGCTGTGCTTTTCAGCC	1353		
QY 1091	tcggccgcgcgcagcagcagcaggtgtcagagatcatcgatggcctctcagagcagag	1150		
DB 1354	ACCATTCTAGGAGTGAGCATGAGATTTCTGAATTTATGATGGGCTCTCTGACCGAGAG	1413		
QY 1151	aacctggaagcagatcgcgccagctggccgtgatcccgcccatctgctgaacgcgtgac	1210		
DB 1414	AAATAATGAAAAAATAATGCGCAGCTCTGTGATTTCCACTATGATGTTTGTATGTCAGAA	1473		







```

RESULT 15
HSU80761/c
LOCUS
DEFINITION Homo sapiens Crc26 alternate open reading frame mRNA, complete cds
ACCESSION U80761
VERSION U80761.1 GI:2565090
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 560)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
cDNAs with long CAG trinucleotide repeats from human brain
Hum. Genet. 100 (1), 114-122 (1997)
2 (bases 1 to 560)
AUTHORS Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
Direct Submission
Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of
Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
FEATURES
Location/Qualifiers
1..560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="fetus"
/tissue_type="brain"
<1..560
/gene="CRC26"
<1..375
/gene="CTG26"
/notes="cysteine rich"
/codon_start=1
/product="Crc26 alternate open reading frame"
/protein_id="AAB91452.1"
/db_xref="GI:2565091"

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 09:38:43 ; Search time 74.42 Seconds  
(without alignments)  
2050.397 Million cell updates/sec

Title: US-09-522-753-5

Perfect score: 13215

Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEPKPLLCQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	13178.5	99.7	2518	21 AAB40574	Human ORFX ORF338
2	7866	59.5	1495	18 AAW18226	Transcriptional co
3	4187	31.7	2453	21 AAB12454	HNRCR protein sequ
4	4043	30.6	2343	21 AAB12453	Human HNRCR protei
5	718	5.4	619	17 AAR99738	Retinoid X recepto
6	565.5	4.3	3266	21 AAB42491	Human ORFX ORF2255
7	477.5	3.6	3118	22 AAB50362	Human SRCAP, Homo
8	469.5	3.6	2819	22 AAB35408	Human 07CG27 gene
9	456	3.5	2971	21 AAB41231	Human ORFX ORF995
10	452	3.4	2972	22 AAB50363	Human SRCAP, Homo
11	429	3.2	2091	21 AAB12000	Rat p3103 protein.

12	428	3.2	1404	13 AAR26049	MSF precursor. Sy
13	428	3.2	1404	22 AAB60568	Human megakaryocyt
14	428	3.2	1404	22 AAB29773	Human megakaryocyt
15	422	3.2	2087	22 AAB31516	Amino acid sequenc
16	400.5	3.0	1464	19 AAW68485	Human recombinant
17	397	3.0	2783	13 AAR23963	Human recombinant
18	394	3.0	2783	13 AAR23962	AFP-1 (Ala 2460 Va
19	382.5	2.9	1012	20 AAY17406	AFP-1. Homo sapie
20	382	2.9	763	18 AAW31852	Human atrophin-1 r
21	381.5	2.9	3910	14 AAR38470	Mycobacterium tube
22	381.5	2.9	3910	16 AAR66462	ALL-1 protein. Ho
23	380	2.9	2286	22 AAB65635	ALL-1 (acute lymph
24	377.5	2.9	1064	17 AAR93254	Novel protein kina
25	377.5	2.9	1064	19 AAW57652	Collagen-like poly
26	377.5	2.9	1065	14 AAR37741	Collagen-like poly
27	377	2.9	1411	21 AAY56800	Human preproalpha
28	367	2.8	1185	20 AAY33497	Human atrophin I p
29	359	2.7	1618	13 AAR27205	Human nestin. Hom
30	358	2.7	1618	15 AAR60127	Human nestin prote
31	356	2.7	3969	15 AAR52971	Product of the CDN
32	353	2.7	1341	16 AAR71701	Collagen alpha 1 (
33	352	2.7	1838	15 AAR53257	Human collagen (Ty
34	351	2.7	1341	21 AAY96122	Collagen type I al
35	350.5	2.7	2742	21 AAR23012	Human APC protein
36	350.5	2.7	2842	15 AAR63508	Adenomatous polypo
37	350.5	2.7	2843	19 AAW76140	Human APC protein
38	350.5	2.7	2843	19 AAW76140	Human APC protein
39	350.5	2.7	2843	21 AAB23011	Human APC protein
40	350.5	2.7	2973	19 AAW76821	Human APC protein.
41	350.5	2.7	2973	21 AAY70304	Protein used in ca
42	350.5	2.7	2973	22 AAY72782	Transcriptional ac
43	348	2.6	3256	21 AAY50976	Human cell cycle p
44	347.5	2.6	2843	16 AAW11922	Adenomatous polypo
45	345	2.6	1388	21 AAY84539	Amino acid sequenc

#### ALIGNMENTS

RESULT 1  
AAB40574  
ID AAB40574 standard; Protein; 2518 AA.  
XX AAB40574;  
XX AC  
XX DT 08-FEB-2001 (first entry)  
XX DE Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.  
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antifungal; antibacterial; antifungal; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX Homo sapiens.  
XX OS  
XX WO200058473-A2.  
XX PN  
XX Human ORFX ORF2255  
XX PD 05-OCT-2000.  
XX PF 31-MAR-2000; 2000WO-US08621.  
XX PR 31-MAR-1999; 99US-0127607.







-Qy 1500 ERACYEESLSKSPGTASSSGGSIARGAPVIVPELKGKPRQSPBLYEDHGAPFAGHLPRGSP 1559  
Db 1501 eracyeeslksrpgtasssggsiargapvivpelkgkprqspblyedhgapfaghlprgsp 1560  
Qy 1560 VTMREPTRELQEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGVS 1619  
Db 1561 vtmreptrelqegslssskasqdrkltstpreiakspstvphehphpispvyehllrgvs 1620  
Qy 1620 GVDLYRSHIPAFDPTSIPIRGIDLDAAYYLPRHLAPNPTYPHLYPPYLIRGYPDTAAL 1679  
Db 1621 gvdlyrshipafdptsipirgidldaaayylprhlapnptyphlyppylirgypdtaal 1680  
Qy 1680 ENRQTIINDYITTSQMHNTATAMAQRADMLRGLSPRESSIALNYAAGRGIDLSQVPH 1739  
Db 1681 enrqtiindyitstqgmhntatamaqramlrglspressialnyaagrgidlsqvp 1740  
Qy 1740 LPVLVPTTGGTPATAMDRLAYLPTAPOPFSSRHSSPLSPGPHLTkPPTTSSSERERD 1799  
Db 1741 lpvlvpttggtpatamdrlaylptapopfssrhssplspgphltkpttsssererd 1800  
Qy 1800 RDRERDREREKSLTSTTTVEHAPVIRPGTEQSSGSSGGGSSSRPASHSHAHQ 1859  
Db 1801 rdrerdreksiltstttvehapvirkpgteqsssgssggssrppashshahq 1860  
Qy 1860 HSPISPTQDALQORPSVLHNTGKGIITAVEPSKPTVLRSTSSVPRPAATPPATHC 1919  
Db 1861 hspisprtqdalqorpsvlhntgmkgiitavepstptvlrststssvprpaatppathc 1920  
Qy 1920 PLGCTLDGVPTLMPEVLLKPEARVAPRPRADTGHAFKAPKPSGLEPASPSPKGS 1979  
Db 1921 plgctldgvptlmpevllkpearvarperpradtghafkappksglepaspaspkgs 1980  
Qy 1980 EPRPLVPVSGHATIAITPAKNLPHASDPDPAPPASADPHREKTSQKPFESIQEELR 2039  
Db 1981 eprplvpvsghatiartpaknlphasdpdpappasasdphektqskpfesiqelr 2040  
Qy 2040 SLGYHGSYSPEGVEPVSPVSSSLTHDKLPHLEELDKSHLECELKPKQPGVKLGE 2099  
Db 2041 slgyhgsyspegvepvspvssslthdklphleeldkshlecelkpkpgvklge 2100  
Qy 2100 AAhLPHRLPLPESOPSSPLLOTAPGVKGHQRVVTLAQHISEVITQDTRHHPQOLSAFL 2159  
Db 2101 aahlphrlplpesopsspllotapgvkgqrvvtlaqhisevitqdytrhhpgqlsapl 2160  
Qy 2160 PAPLYSPFGASCPVLDLRRPDSILYLPDPDHGAPARGSPHSEGGKRGPEPKTSVLGGGE 2219  
Db 2161 paplyspfgascpvlldlrrpsdilylppdhgaparpsphseggkrpsepktsvlggge 2220  
Qy 2220 DGIIEPVSPPEGMTEPHGHSRAVYPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESN 2279  
Db 2221 dgiiepvsppegmtpeghrsavypillyrdgeqtepsrmgskspgntsqppaffskltesn 2280  
Qy 2280 SAMVSKKQELINKLNTHNRNEPEYNTSQCTEIFNPAITGTGLMYRSQAVQEHASTN 2339  
Db 2281 samvskkqelinklnchnrnepeynlsqgpteifnmpaitgtglmyrsqavqehastn 2340  
Qy 2340 MGLBAIIRKALMGKYDQWESPPLSANAFNPLNASILPAAMPITAADGRSDHPLTSPGG 2399  
Db 2341 mgleaiirkalmgkydqwespplsanafnplnasaslpampitaadgrsdhpltspgg 2400  
Qy 2400 GGAkakvSGRPSRRKAKSPACGLAGDRPPSVSVHSGDCNRRRPLFNVRWEDRPSAGS 2459  
Db 2401 ggaakvsgprsrkakspaglagdrppsvsvshsgdcnrrrplfnvrwedrpssags 2460  
Qy 2460 TFPYPNPLMRLOAGVNASPPPCGLPAGSPLAGPHAWDEEPKPLCSQYETILSDSE 2517  
Db 2461 tpfypnplmrloagvmasppppgipagsgplagphawdeekpplcsqyetilsdse 2518

RESULT 2

AAW18226

ID AAW18226 standard; Protein; 1495 AA.

XX

AC AAW18226;  
XX 24-SEP-1997 (first entry)  
XX Transcriptional co-repressor SMRT.  
XX Silencing mediator for retinoic acid and thyroid hormone receptor;  
KW SMRT; transcriptional co-repressor.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
XX Region 1..160  
FT /label= N-terminal\_region  
FT /note= "proline-rich domain"  
FT Region 773..790  
FT /label= ERDR\_region  
FT Region 812..827  
FT /label= SG\_region  
FT Region 1061..1132  
FT /label= glutamine-rich region  
FT Region 1201..1495  
FT /label= C-terminal\_region  
FT Peptide 1330..1376  
FT /note= "alternatively spliced insert not present in the original two-hybrid clone"  
XX  
PN W09709418-A1.  
XX 13-MAR-1997.  
XX 24-JUL-1996; 96WO-US12371.  
XX 01-SEP-1995; 95US-0522726.  
XX (SALK ) SALK INST BIOLOGICAL STUDIES.  
XX Chen JD, Evans RM;  
XX WPI; 1997-192894/17.  
XX  
XX New co-suppressor of steroid-thyroid hormone receptor activity -  
PT also methods for identifying compounds that relieve its suppressant  
PT effect and/or activate receptors  
XX  
PS Claim 2; Page 40-45; 71pp; English.  
XX  
CC A novel receptor interacting factor (AAW18226) is designated SMRT,  
CC i.e. silencing mediator (co-repressor) for retinoic acid receptor  
CC (RAR) and thyroid hormone receptor (TR). Its association with RAR  
CC and TR both in solution and on DNA response elements is  
CC destabilised by ligand. The interaction of SMRT with mutant  
CC receptors correlates with the transcriptional silencing activities  
CC of receptors. In vivo, SMRT-functions as a potent co-repressor. A  
CC GAL4 DNA binding domain fusion of SMRT behaves as a frank repressor  
CC of a GAL4-dependent reporter. These data identify a novel class of  
CC cofactor which is believed to represent an important mediator of  
CC hormone action. Full-length cDNA for SMRT has been isolated from  
CC a HeLa library in a two-hybrid screen using a GAL4 DNA binding  
CC domain/RXR fusion protein as bait.  
XX  
SQ Sequence 1495 AA;

Query Match 59.5%; Score 7866; DB 18; Length 1495;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1486; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1031 DKEAFAAEAKLPGDPCWTSGLPFPVPPPREVIKASHADPPSAFSAFGCHPPLGLHD 1090  
Db 9 dkeafaaeaqklpgdpcwtsglpfpvppprevikasphadppsafsfyapghpplglhd 68











Db 2021 estrgyegp-----lhhyrsqespspqgqplppssqsegmgqvrthrliatd 2071  
QY 2138 HISEVITQDYTRHH-PQOLASAPL-----DAPLYSFCGASCPLVDLRRPPSDLYLPPPD--- 2189  
Db 2072 hicqitqdfanqpsqaststftspalastp-----vrtktsrysesqsgt 2123  
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGEGDIEPVPSPGEMTEPG-HSRSA 2240  
Db 2124 vlhprpgprvspenlvdkargprgkspersh---pspeyepispqg---pavhekqd 2177  
QY 2241 VYPLLYRDEQTEPERMSKSCPNTSQPPAFSKLTESAMVKSKEIKKMLNTHNRN 2300  
Db 2178 smlllsgrvpaedgrsdrrspsyslpfftkl-estspmvkskqefrklinsggg 2236  
QY 2301 EPEYNISQGTIFNNPATGTGLMTYRQAOVQEHASTNMGLEAIRKALMGKYDQWEE- 2359  
Db 2237 dsdmaaaqgtetifnlpavttsgavssrshsfadpas-nigledlrkalmgfsfdkved 2295  
QY 2360 -----SPPLSANAFNLNASASLPAAMPITAADGRSDHDLTSPGGG-GKAKYSGRPSRRK 2413  
Db 2296 hgvvmshpv---gimpgaastsv-----vtssearrdegepsphagvckpklinsnrk 2347  
QY 2414 AKSPAPGLA--SGDRPPSVSVHSGDCNRRPPLNVRWEDRPPSAGSTPPPYNPLIMRL 2471  
Db 2348 skspipggsylgterpsvsvhsegdyhrqtp--gwawedrpsstgstqfpyngltirm 2405  
QY 2472 QAGVNASPPPPGLPAGSGPL--AGPH--HAWDEPKPLLCQVETLSDSE 2517  
Db 2406 ----lsstpttiacapsaitqaahqgnriwerepapllsayaqetclsd 2452

## RESULT 4

AA12453 ID AA12453 standard; Protein; 2343 AA.

AC AA12453;

XX 24-OCT-2000 (first entry)

XX Human HNRCR protein SEQ ID NO:20.

XX Human; HNRCR; nuclear receptor coreceptor.

XX Homo sapiens.

XX CN1250094-A.

XX 12-APR-2000.

XX 06-OCT-1998; 98CN-0120919.

XX 06-OCT-1998; 98CN-0120919.

XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.

XX Yu L, Tu Q, Zhao Y;

XX WPI: 2000-400830/35.

XX N-PSDB; AAA60829.

XX Preparation of new human keron acceptor co-repressor coding series and

XX the polypeptide -

XX Claim 4; Page 25-30; 58pp; Chinese.

XX The present sequence represents a human homologue of nuclear receptor

XX coreceptor (HNRCR).

XX Sequence 2343 AA;

Query Match 30.6%; Score 4043; DB 21; Length 2343;

Best Local Similarity 41.3%; Pred. No. 2.1e-219;

Matches 1061; Conservative 330; Mismatches 731; Indels 444; Gaps 104;

QY 137 KDRSLTGLKLE-PVSPSPPHDTDPELELVPPRLSKBELQNMDRVREITMTWEOQISKLLK 195

Db 36 qdpafggkheapspsisgqpcgddqnaspsksleelqsgmdrvdrelakveqqlklkk 95

QY 196 KQQLLEEAAPPEPEKVPKVPPIESKHSRLVQIIYIDNRKKAEEAHLLEGLEGPQVELP 255

Db 96 kqqqlleeeaaapepekpvsppveqkhrsvlqvliidenrkkaeeahkfeglgpkvelp 155

QY 256 LYNQSDTROTHENIKINQAMRKKLILYFKRNHARKQWKQFCORYDQLEALEKKVVER 315

Db 156 lypqsdtkvynheniktnqvmrkkilifkrnrharkredqkicdydqlmeawekvdr 215

QY 316 IENNPARRAKESKVREYKOFPEIRKQRELQRMQSRVGORGSLSMSAARSEHEVSEI 375

Db 216 iennprrrakesktrreyekqfpeirkqregqrf-qvgrgaglsatarseheisei 274

QY 376 IDGLSEQENLEKQMRQLAVIPPMYLDADQORIKFINNGLMADPMKYKDRQVNMWSEQ 435

Db 275 idglsegesnekqmrqlsvippmmfdaeqrkvkfimnglmedpmkvykdrqfmvwtidh 334

QY 436 EKETPREKFMQHPKNGFLIASFLERKTVAECVLYIYLLKKNENYKSLVRSY-RRRGSQ 494

Db 335 ekeifkdkfikhpkngfliasylerksvpcvlyylytkknenykalvrnygkrgrnq 394

QY 495 QOQOQOQOQOQOQOQOQPMRPSOEEKDEKEKEKEEKEPEVENDEDLLEKTDOTS 554

Db 395 qlarpsqeekeveeke--dkaektkekkeekdeekedekeskentke--kdkidgta 449

QY 555 GEDNDEKEAVASKGRKTANSQGRRRGRITRSMANEEAEITPQOS-----AEL 604

Db 450 -eeteereqatprgrktansggrkgritrsmtneaaaaaataateeppppppppep 508

QY 605 ASMELNESSRWTEEMETAKKGLLEHGRNWSAIAIRWGSKTVSQCKNFFYFKKQNLDE 664

Db 509 lstepvetsrteeemevakkglvehgrnwaaiaakmvgtkseaqcknfyfnykrhnlidn 568

QY 665 ILQOHLKMEKERNARRKKKAPAAASEEAFFPPVVEDEMEASGVSGNEEMVEAEAL 724

Db 569 llqhqkqktsrpreerdvqcesvastvsa---qededleas---neenpedseg- 619

QY 725 HASGNEVPRGECGPATVNNSSDTEIPSPH-TEAAKTQNGQPKPATLGAAGDPPPPGP 783

Db 620 -----aenssdtesapspsveavk-----psed 643

QY 784 TPRRTSRAPTEPTPASEATGATPPAP-PSPS-APPVVPVPEKEKEBETAAPV---- 837

Db 644 spenatsrgntepavele-----pttetapstpslavpstkpaedesvetqvndisaet 699

QY 838 -----BGEEOQKPPAAEELAVDTGKAEEPVKSECTEAEAGGPAKGDAAEAE 884

Db 700 aeqmdvdqgehsaeeegsvcdppppack--adsvdvevrvpnenhaskvgedntkerldra- 756

QY 885 ATAEGALKAEXKEGSGRATTAKSSGA-----PQ-DSDSATCSADEVDEAEGDKNRLLS 939

Db 757 -----sekveprdedlvvaqinaqrpeqpsndssatscade--dvdgepergrmf 806

QY 940 P---RPSLLTPTGDPANASPOK--PLDLKQLKQRAAAIIPPI----- 976

Db 807 pmdskpsllnptgslv--sslikpnpldlpqqlghraavippmvscstpcnpiptvpsvga 865

QY 977 ---QVTKVHEPPREDAAPTKPAPPAPPPQNLOPESDAPQOQSSPRGKSRSPAPPADK 1032

Db 866 lyqrhikamhesalle-----eqqrqeqidlecrrstscpgtsksp-----nr 909

QY 1033 EAFAAEAQKLPDGPCTWTSGLPFPVPPREVIKASHPDPSPAFSAPPCHPLPLGLHDTA 1092

Db 910 e-----w-----evlqaph-----gvitnlpegvrlp-----tt 934

QY 1093 RPVLPRPPTISNPPPLISSAKHPSVLRIQIGAISOG-----MSVOLHVYPYSEHAKAP-VG 1146

Db 935 rtrpppppplpskttvasak-psi--mgsisicgtptgtitshnqasytqetpkpsvg 991







```

SQ Sequence 619 AA;
Query Match 5.4%; Score 718; DB 17; Length 619;
Best Local Similarity 36.0%; Pred. No. 1.5e-32;
Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;
QY 1975 PSKGEPRPLV-----PPVSGHATIARTPAKN--LAPHASDPDPAPPASADP 2021
Db 70 psgkqphasvvyseagdkgppksryeeelrtgrkttitaanfdivttrqiasdkda 129
QY 2022 HREKTSKPFISQIELESLRSLGVHGSYSPEGVPEPVSPPSSSLTHDKGLPKHLEELDKSH 2081
Db 130 rergsqsdss-----sslsrhryetaadaievispassappqekpqayqpdmvkanq 183
QY 2082 LGEURPKQPGPVKLGGEAAHPLHUR-----PLPSQSSSSPLLQT--APGVKGHRVVT 2134
Db 184 aenestrgyegp-----lhvrsqdespsqgqplpsssqsegmgvprthrilit 234
QY 2135 LAQHSITSEVITDTRHHPOQLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189
Db 235 ladhicqitqfarn--qvpqststftqpsalsstcpv---rtktsryspesqst 289
QY 2190 --HGAPA--RGSPH-----SEGKRPSPNKTSVLGGGEGIEPVSPPGWTPEFG--HSRSA 2240
Db 290 vlhprpgrvrspenlvdkrsrgspkspershi---psepyepispqg---pavhekqd 343
QY 2241 VYPLLYRGEQTEPRMGSKSPGNTSQPPAFPSKLTESNAMSVMKSKOEINKKLNTHNRN 2300
Db 344 smllsqrgvdpaedrsrpsgsisylpsfftkl--estspmvksskqefrklnssggg 402
QY 2301 EPEYNISOPGTIFNMPLTGTGLTYRQSOAEOEHASTNMGLEATIRKALMGKYDQWEE- 2359
Db 403 dsdmaagpqtfeiflvpavttsgvssrshsfadpas--nigldlirkalmgfsfdkved 461
QY 2360 -----SPPLSANAFNPLNASALPAMPITAADGRSDHTLTPSGGG--GKAKYGRSPSRK 2413
Db 462 hgvvmshpv--glnpgsaustav-----vtssearrdegepsphagvckpklinksnrk 513
QY 2414 AKSPAPGLA--SGDRPPSVSVSHSGDCNRRTPLNLRVWEDPSSAGSTPPFYNPLMLRL 2471
Db 514 skspipgsgylgterpsvsvshsgdyhrqtp--gwawedpsrsgtqfnpitirm 571
QY 2472 QAGVMASPPPPGLPAGSGPL--AGPH--HAWDEEPKPLCQYETLSDSE 2517
Db 572 ----Lssptqlacapsaitqaqhqnriwerrepapllsaqyetlsdsd 618

RESULT 6
AAB42491
ID AAB42491 standard; Protein; 3266 AA.
XX AC AAB42491;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2255 polypeptide sequence SEQ ID NO:4510.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX KW vulnary; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX KW antitubercular; antifungal; antirheumatic; antihypertensive;
XX KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX KW cardiovascular disease; diabetes mellitus; erythroidism; SCID; AIDS;
XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX KW thrombosis; contraceptive.
```



Db 942 kqagrfdvspnsii-----krdslrksvrdl--epgevpsdsdedghekskshsprasal 995  
Qy 323 -----RAKESKREYVEYKOPPEIRKORLEQWQSRVQORGSL-----SMSAAR 367  
Db 996 yessrlsflldredkrlenderlssierknkfyfaldktitpdtkalleraksissr 1055  
Qy 368 SHVEVSEIIDGLSE-----OENLEKOMROLAVIPPMYLDADOOIRKIFINMGLAMPKMY 423  
Db 1056 een--wsfldwsrfanfnnkdkvdsapripwsymkklir--tdsegkmdkdkdh 1112  
Qy 424 KDRQVNMWMSOEKTFREKFMHPKNGLTASFLEKRTVAECVLYLYLTKNENYKSLV 483  
Db 1113 keee-----qergelfasrlf-hsifeqskrlq-----hlerkeedsdfis 1154  
Qy 484 RRSYRRRGKSQQQQQQQQQQQQQQOQOQP-----MPRSSQBEKDEKEKEKEAEE 533  
Db 1155 griy--gk-----qtsegansttdsiqepvlfhsrfmeltmqgkkekdkpkeveq 1208  
Qy 534 EKPEVNDKEDLLKKTDDTSGEDNDEKEAVASKGRKRTANSOGRKRGRITRSMANEANSE 593  
Db 1209 ---dten-----hpktpesapenk-----selktp 1231  
Qy 594 EAITPQQAELASLMELNESSRWEEMETAKGLLEHGRNMSAIARMVSGKTVSOCKNFY 653  
Db 1232 psvgpv-----tvvtlesap-----salektgdktv----- 1260  
Qy 654 FNYKKRONDELTOOHLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGVSN 713  
Db 1261 -----eaplvteektvepatvseeakpasepapa 1289  
Qy 714 EEMVVEEAELAHAGNEVPRGECGPATVNNSSDTPESPHTEAAKTGQNGPKPPATL 773  
Db 1290 pveqleqv-----dippga-----dpckeaaamp--agveegsggqpp-yl 1328  
Qy 774 GADGPPGPPPTPPRTSRAP--IEPTPASEATGATPPAPPPSPAPPPVPVPEKEKEEET 831  
Db 1329 dak-----pptpgasfsgaenvdpepds-----tqlskpaqkseaneakpdpdat 1377  
Qy 832 AAAPP-----VEEGEORPPAAEELAVD---TGKAEPEPVKSECTEEAEPEGKAGKDAEAA 883  
Db 1378 adaepdqakeaapesppasedlevdpvaaakdkkpnkkskrkrtpvqaaavsvivekpv 1437  
Qy 884 EATBEGALKAKEKGGGRATTAKSSGAPQSDS--SATCSADEVDEAEAGDKNRLLSRP 942  
Db 1438 trkseridreklnksprgeaqllelmeaekitrtasknsaadle-----hpep 1489  
Qy 943 SL-LTPT-----GPRANASPK-PLD-----LKQKQRAAAIPPIQYTK 980  
Db 1490 slpstrrrrnrvsvyatmgd--henrsvpkepvqprvtrkrlerelqeaavp--ttpr 1546  
Qy 981 VHEPPR-----BDAAPTTPAPPAPP-----PQNL-----QPESDAP 1012  
Db 1547 rgrpbttrradeeeneakepaetlkpegwrsprsgktaaggpggqgkknepkvdad 1606  
Qy 1013 QOPGSSPRG-----KSRGPAPPADKFAEAQKLPDPGPPCWTGSLPFPVPPPREVIKASP 1067  
Db 1607 rpeattevgpqlgvkessmepkaaeaeagseqkdrkd--agtdknnpbetapvvevekkp 1664  
Qy 1068 HAPDPSAFSYAPPHPPLGLLHDTPARVLPRLPPTTSNPPPLISSAKHFSVLEROIGALISO 1127  
Db 1665 -apeknks-----krgsrnsrlavdkksa1kn-----vdaavsprgaaagageres 1711  
Qy 1128 GMSVOLHVPYSEHAKAPGVPTMGLPLMDP-----KKLAPPSG-----VKOEOL-- 1172  
Db 1712 gy-vavspkesesqpkedglssqlksdpdpokekedvsaasgspseatqlakgmeleq 1770  
Qy 1173 -----SPRGAGGPPELSGVPTAQEAASVLRGTALGSV----- 1203  
Db 1771 avehlaaklaeasasaaykadapegla--pedrtdkphahasetetelaaalgsindisgepe 1828  
Qy 1204 -----PGGSITKIPSTRV-----PSDSAITYRGSTTHGTPADVLYKGTI-TRIGIED 1250  
Db 1829 nfpappppypgesqldlqpagaqalqpse-----egmetdeavsvsiletteaates 1878

Qy 1251 S-----PSRLDRGRDSDSLPKGHVIEYEGKKGHVLSYEGGMSVTQCSKEDGRSS 1298  
Db 1879 srppvnadpsagpdtckeargnsetshtsvpeakgsk-----eventlvrkdkgrq-- 1930  
Qy 1299 GPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPERHSPHHLKEQHHRG---SI 1355  
Db 1931 -----ktrsrkrntnkkv-----vapveshvp-----esndaqagespaa 1966  
Qy 1356 TOGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTOALGPLKLKPAHEGLV 1415  
Db 1967 negttlvqpeapq-----eekqsexphstppscdskskipste----- 2006  
Qy 1416 ATVKEAGRSIHIEPRELRHT-----PELPLAPRLKEGSITQGTPLKYDYDTGASTGSKK 1470  
Db 2007 -----nssqseisveertptkasvpdpdlppppqp-----apvdeepqa-----r 2044  
Qy 1471 HDVRSLSIGPGRTFP--PVHPLDVMDARALERACYEESLKSRRPGTASSGSGSIARGAPV 1528  
Db 2045 frvhsliesdpvtppsdpsipptlpsv-----taaklspvvasg-- 2084  
Qy 1529 IVPGLGKPROSP-----LTVEDHGAPFAGHLPRGSPVTMREPTRLQEGSLSS---SK 1578  
Db 2085 -----gibqspptkvwtitrqe-----epraqstpspalpdpdkasdvds 2127  
Qy 1579 ASQDRKLSTPREIAKS-----PHSTVPEHHPHPISPYEHLRLRGVSGVDLYR 1625  
Db 2128 estllkilmdpkvysaatsvtstsvtataeapvsaapclheappppvd-----skkplee 2181  
Qy 1626 SHIPLAFDPTSIPIRGIPLDAAAAYYLPHLAPNPIYPHYLYRIGYDPTAALENRQTI 1685  
Db 2182 ktappvtnnseiqasevivaadkekvaplapkit-----svismrmpvsidlensqki 2234  
Qy 1686 INDYITSQOMHINTATMAQAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745  
Db 2235 -----tlakpaptltgl-----vsaltglvnlsvlp-vnalkg 2267  
Qy 1746 PTPGTPATAMDRLAYLTAPOQPFSSRRHSSPLSPGPGTHLTPTTTSSSERERDRDRD 1805  
Db 2268 pvkgsvtllkslvs-----tpagpvnlkgpv----- 2294  
Qy 1806 RDREKESILTSTTVEHAPI-----WRPGTQSSGS-----SGSS 1841  
Db 2295 -----nvltpgvnltpvnaatvgtnaapgtvnaaasavnatasaavtvtagavtaas 2347  
Qy 1842 GGGGS-----SSRPASHSHAHQHSPTSPRTQDALQORPSVLHNTGMKIIT 1888  
Db 2348 ggvattgtvtmagaviapstckckgrasanensrflpagsmpvldirpa---dag-sgaql 2403  
Qy 1889 AVEPSKPTVLRSTS---TSSSPVRPAA---TFPPATHCPLGGTLDGVYPTLMEPVLPLKEA 1942  
Db 2404 rvntsegvvllysygqktegpgqrisaklsqilppas-----amdiefgqsvaksqvkpds 2457  
Qy 1943 PKVARP--ERPRADTGHAPLA-----KPPARSGLPEAPSSPSKSGSEPRPL--- 1984  
Db 2458 vtasqpskpgqpagyanvathstlvtltaqtnaspvissvk-adrpal-ekpepilis 2515  
Qy 1985 -----VPPVSGHATIAITPA----- 1999  
Db 2516 vstpytqggtkvltcggintppvlvhnqlvltpsvlttnkkladpvtlkietkviqpanl 2575  
Qy 2000 -KNLAPHUASPPPPAPPASADP--HREKTQSKPF--SIOELELRSLGVHSGSYSPGEGVE 2054  
Db 2576 gstlcpvh-----ppalpsklptevnhvpsgspisadrtvshlaaakldahsprpgp 2631  
Qy 2055 PVSPVSPSLTHDKGL-----PKH--LEELDKSHL- 2082  
Db 2632 sfrashpsstastalstnatvmlaagipvpqfissihpeqsvimpphsaitqvtvslshls 2691  
Qy 2083 EGELRPKOPG-----PVKLGGEAAHPLPHLPLPESQSSPPLQOTAPGVKGHORVVTLAQH 2138  
Db 2692 ggevrmntptlpsitcitysirpealhspr-apl---qp-----qqievra--- 2730







Qy	663	DEILQQH-----	-----KLKMEKNARRKK	683
Dz	743	phvfqshrefkewfsnplgmiegsgenyeglvkrhlkvrlprfllrrvkdvemqpkky	802	
Qy	684	K---KAPAAASEEAAPPVVVEDEEMASCVSGNEEMEVEAEALHASGNEVPRCECSGA	740	
Dz	803	ehvlrcslsrqrclayddfmaqtktketlatghfmvsnilmql-----rkvcnhp-	853	
Qy	741	TVNNSDTEISPP-----	HTEAAKDGTONGPKPATLGCAGD---	777
Dz	854	--nlfdprvtspfiftpgicfstaslvratdvhplqridmgrid-----liglegrvs	905	
Qy	778	-----pppgp-----	PTP-----	PR 787
Dz	906	ryeadtflphrisrvillevatapdpbrpxkmkvnmrlqpvrkqegrtvvvvnnpr	965	
Qy	788	R-TSRAPIETPASETGAPTTPAP-----	PSPSAPPVPVVKKEEETAAP	836
Dz	966	apiqpvvprrppgeiselagtpqvqvilpaslmvsasp-aggpilip-asrrpgpavlilpp	1023	
Qy	837	VEEGEQKPPA-AEELAVDTGKAEPVKSECTEEAECPKAGKDAEAAATAEGALKAEK	895	
Dz	1024	lqpnsgslpqvlpsplgvlsgrtpptclsiktpc--papvrlspapppgssallkplt	1081	
Qy	896	KEGSGRATTAKSGAPOSDSSATCSADEVADEBGGDKNLLSPRLSLTPTGD----	950	
Dz	1082	vppg---ytfppaaattstttatattavpaptpaqriilspdmqarlpsevvsvig	1137	
Qy	951	-----PRANASPOKPIDLK-----	QUKQ-----RAAIPPIQVTK-----	980
Dz	1138	qlasladrpanaggskpfiflggnkkitltgaqvrqlavggprlqmptmvnntgvk	1197	
Qy	981	--VHEPPREDAATKPAKPAPPQPQNLPESDAQO---OPGSSPR---	GKRSRAP-----	1028
Dz	1198	ivvrqardgltpvvpilapaparppssgilpavlnprrtltpgrltpctigtarapmtcptl	1257	
Qy	1029	-----PADKEAFAAEAOCLKDPDCWTGSLFPFVP-----	PREVIKASPHA	1069
Dz	1258	vrpllkivhspsvevsasa----pgaapl-tissaplhvpslpgpaassmpipnsspla	1311	
Qy	1070	PDSASFAYAPGHPLGLHDHARVLRPPPTIISNPPLISSAKHPSVLERGIAISQGM	1129	
Dz	1312	spvsstsvplsslslpisvptllpapasaapltipisapltvsaagpal-----	1360	
Qy	1130	SVQLHVYPYSEHAKAPGVYTMGLPLMPDPKKLAIFSGVKQEOLSPRCOAGPPES---	LGV 1186	
Dz	1361	-----tsvtcpilapvyvaapgp-----	pslqpsg-apspsasaltigl	1396
Qy	1187	PTAQ--EASVLRGTALSGVPGGSITKGITPSTRVPSDSAITVRGSTTHG-----	1232	
Dz	1397	atapslssqtpqbhlillaptashvvginstvapacsylvpasalasfpasnppapaq	1456	
Qy	1233	-----TPADVLYKGTIRIIGEDSPSRDLRGREDLSLPKHGVITY----	ECKKHVLSYEGGM	1284
Dz	1457	asllapasasqalatplapmapaqdtailapsapaplaplpvlapsppaapvlassctp	1516	
Qy	1285	SVTCCKEDGRS--SSGP--PHETAAPKRYIDMMEGRVGRAI--SSASIEGL-MGRAIPP	1337	
Dz	1517	pvmapstptgtslasasvpaptpvylapsstqtmipavpvspilpspastgtlalalalap	1576	
Qy	1338	ERHSPhLKQHHRIGSITOGIPRSVVEAQEDYLRRACKLLKREGTTPPPPSSRDLEAY	1397	
Dz	1577	t-----lqgs-----	-----spstqlslgtgnqqgfptqtlls---	1604
Qy	1398	KTOALGPLKLKPAHEGLVATVTKAGRSIHETPRELRTPELPLAPR-PLKEGSITOGTP	1456	
Dz	1605	-----ltpa-salvpt-----	paqtislapgpplgtqlslalapappilapsapvpap	1651
Qy	1457	LKYDTGASTTGSKK-----HDVRSLLGSGPRTFPVPHLDVNMADARALERACYEESLKSRP	1512	
Dz	1652	ahtltlapasssasllapasavtqltlspa-----	pvtligpaaaqtlalapastqspa	1704

QY	1513	GTASSGGGIAGAPVIV	-----PELGKPRQSPQLTYEDHGAPAGHLPRGSPV	1563
Db	1705	sqasslvvsasgaaplvmvslrpskdepdltlrlsgppspstatsfggprlrqp	-----GSLSSSKASQDRKLTSTPREIA	1763
QY	1561	THRETPPR	-----LQE	1593
Db	1764	-----ppprsflyldsleekkrqrserlerifqlseahgalapvygtevlfdctllpqpva	-----HPPHPISPYEHLLRGVGDLYRSHRIPLAFDPTSPRGI	1820
QY	1594	K--SPHSTVPEH	-----HPHPISPYEHLLRGVGDLYRSHRIPLAFDPTSPRGI	1641
Db	1821	spigrpspsbptfwtysteahhravlfqqrldqlseierlifvmp	-----pvea	1872
QY	1642	PLDAAAAYLPHRAP	-----NPYPHLYPPYLI	1675
Db	1873	pppsihachppwlapraqafqeqlaselwprarphrivrvcnmrtfpdllrllydcgkl	-----	1932
QY	1676	TAALLENRO	-----TIINDYITSO	1712
Db	1933	qtlavllrllgkaeghrvllftqmtmlwdieqfityghlyrlldgstrvegrqaimerf	-----QMHHTNATAMAORADMLRG	1992
QY	1713	-----LSPRESSLALNAAAGRGIIDISQVPHLPVLVPPTGCTATAMADR	-----	1757
Db	1993	nadkrifcflstrsgvgvnltagdvvfysdw	-----nptmdaaqadchrigr	2043
QY	1758	-----	-----LAYLPTA	1770
Db	2044	qtrdvhlyrlisertveenilkkanqkrmldmaieggnfrraykqqtirelfdmplee	-----	2103
QY	1771	RHSSS-PLSP	-----CGPHLHKPTTSSSERDR	1800
Db	2104	psssvpsapeeeetvasqkthileqalcraeedeeditraatkaeqvaelaefnendg	-----	2163
QY	1801	-----DRDRDREREKSILTSTTVEH	-----	1823
Db	2164	fpagegeagrpgadeemsraeqaelaaveqltpieryamkfleasleevsreelikqae	-----	2223
QY	1824	-----APIWR-PGTEQSSGSSG	-----SSGGGGSSSRPASHSHAHQH--SPIS	1864
Db	2224	eqveaarkldqakeevfrlpqeeespgagdesscgtggthrrskakapergtvrs	-----	2283
QY	1865	PTQDALQRPVSLHNTGMKGIIITAVEPSKPTVLRSTSI	-----SSVPRPAATFPFATHCPL	1921
Db	2284	erlrgaraetqganhtp	-----visahq	2333
QY	1922	GGTLDGVYPTLMEPVLPLKEAPRVARPERPRADTGHAFKAPRGSGLEPASSPSKSGSEP	-----	1981
Db	2334	pasapaapal	-----vvpvysapvpsiapnptilpvhilpspppsqilpcsspa	2388
QY	1982	RPLVPVSGHATARTPAKNLAPHHASP	-----DPPAPPASAS	2030
Db	2389	ppactpppahtp	-----ppactcivtpssplllgppsypisavtnlplgrpeaelcaqal	2445
QY	2031	FSIOELELSLGHGSS	-----YSPEGVEPVS	2072
Db	2446	aspelelasvassetsslslyppkdlpavellpvsseknlsitpsapsiltleagisn	-----SSPSLTHDKG-LPK	2505
QY	2073	HLEELDKSHLEGELRKPQPGVKLGGEAAHLP	-----HLRPLUPESQSSPLLQTAGV	2126
Db	2506	gqeqeapdaaeattitvlp	-----egeelpcvseenglelppsaadeplqeplead	2558
QY	2127	KGHRQVVTIAQHISEVITODYTRHHPPQOL	-----SAPLPAPLYSFPFGASCPLVLDLRPPSDLYL	2185
Db	2559	r-----tseelteaktptsspekplvtvaeavapstssatssp	-----	2598
QY	2186	PPPDHGAPARGSPHSGGRKSPFNKTSVLGGGEDGIEPVSPPGEGMTEPGHSRSAYVPLL	-----	2245
Db	2599	-----	-----eggsparpr	2608
QY	2246	YRDGEQTEPSRMGSKSPNGTNSOPPA	-----FFSKLTESNAMYKSKQOEINKKLTNHRNEPE	2303
Db	2609	rttsadvelrgggtgrgp	-----qppgpkvrlrkpgrlvtvvekelvqrrr	2655
QY	2304	YNISOPGTEIFNNPAITGTGLMTYRSQAOVQEHASTNNGLEAIIRKALMGKYDQWEESPPL	-----	2363















QY 1235 ADVLYKGTITRIIGEDSPRLDRGREDLSLPKGVHVIY-----EGKKGHVLSYEGGMSVTOCS 1290  
Db 1317 aasasqalatlapmaaaqatallapsappaplaplpvapsagaapvlassqtpvpmvmaps 1376  
QY 1291 KEDGRS---SSGP---PHETAAPKRYDMMEGRVGRAI--SSASIEGL-MGRAIPIPERHSPH 1343  
Db 1377 stpgtstlasaspaptpvlpasstqtmldpavpslpaspastqtlalapalapt----- 1431  
QY 1344 HLKEQHHRGISTOIGIPRSYVEAOEDYLREAKLLKRGCTPPPPPSRDLTEAYKTQALG 1403  
Db 1432 -----199s-----spstqlslgtnpgqppftqtqls----- 1458  
QY 1404 PLKLPAPHEGLVATYKEAGRSITHEIPRELHRTPELPLAPR--PLKEGSTTOGTPLKYDTG 1462  
Db 1459 ---lcpa-sslvpt---paqtlslapgplgptqlslapapplapaspavpgpapahtitl 1511  
QY 1463 ASTTSGSK-----HDVRSLIGSGRTFPVPHPLDVMDARALERACYEESLKSRTGATSS 1518  
Db 1512 apasssasllapasvqtlslpa-----pvptlqpaqaqlalalastgspasqassl 1564  
QY 1519 GGSIRAGAPVIV-----PELCKPRQSPLTYEDHGAPFAGHLPRGSPVTRREPT 1566  
Db 1565 vvasagaaplvvmvmsrlpvsksdepdtlrlsgppspststatsfgprprqp-----ppp 1620  
QY 1567 PR-----LQP-----GSLSSSKASQDRKLTSTPREIAK--SPH 1597  
Db 1621 prspfylsleekrrqrserlerifqlseahgalapvygtevldftclpqvaspiapr 1680  
QY 1598 STVPEH-----HPHPTSPYHLLRGVGVLDYRSHIPLAFDPTSIPIGIFLDA 1647  
Db 1681 spgshptfwtteahravlfpqrdqlseierfifvmp-----pveapppslh 1732  
QY 1648 AYLPRLHAP-----NPTYPHLYPPYLI-----RGYPD-----TAAL 1680  
Db 1733 achppwlapraqafqeqlaselwpraprlhrivcnmtqfplrlgdydcqlqlavl 1792  
QY 1681 NRQ-----TIINDYITSQ-----QMHHTATAMAOQDMLRG----- 1712  
Db 1793 lqqlkaeghrvltfqtmdrlvleqfityhghlylridgstveqrqalmerfnadkri 1852  
QY 1713 ----LSPRESSLALNYAGPRGIIDLQVPHLPVLVPTPTGPATAMDR----- 1757  
Db 1853 fcfllistrsgvgvnlgtadtvvfydsdw-----nptmdaqaqdrchrigqtrdvh 1903  
QY 1758 -----LAVLPTA-----PQPFSRRSSS- 1775  
Db 1904 lyrlisertveenilkkanqkrmldmaiegnfnftayfkqqtirelfdmpleepsssv 1963  
QY 1776 PLSP-----GGPHTLTKTPTTSSSERDR----- 1800  
Db 1964 psapeeeetvaskthileqalcradeediraatqakaeqvaelaefnendgfpag 2023  
QY 1801 -----DRDRDREREKILSTTTVEH----- 1823  
Db 2024 eeagrpgaedeemsraeqeialaavleqlptieryamkfleasleevsrealkqaeqveaa 2083  
QY 1824 -----APTWR-PGTQSGSSG-----SSGGGGSSSPASHAHQH--SPISPTQDA 1870  
Db 2084 rkdldqaeevfrlpqeeegpgagdesccgtggthrrskakaperpgtrvserlrga 2143  
QY 1871 LQQRPSVLHNTGMKGIITAVPSPKPTVLRLSTST-----SPVPRPAATFPATHCPLGLTLDG 1927  
Db 2144 raetganhtp-----visahq-----trsttprcparervrpraprptpasapa 2193  
QY 1928 VYPTLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASPSKSGSEPLVPP 1987  
Db 2194 alpal---vpvpsavpvisanpnitilpvhilpspppsqipccspa--ctppactp 2248  
QY 1988 VSGHATIARTPAKNLAPHASP---DPPAPPASAS-----DPHREKTSQKPSIQEL 2036  
Db 2249 ppahtp---ppaqtclvtppspillgppsvpissavtnlplglrpeaelcaqalaspsel 2305

QY 2037 ELRSLGYHGSS-----YSPGEGVPVS-----pv-----SSPSLTHDKG-LPKHLEELD 2078  
Db 2306 elavasasetslsvppkdlipvaveilpvseknsltpsapstleagslngqeqea 2365  
QY 2079 KSHLGEERLPKQPVKVLGGEAHLIP-----HLRPLPESQSPSSSPLLTQAFVKGHQRV 2132  
Db 2366 pdaegttltvlp-----egeelplcvsesngielppsaasdeplqpleadr----- 2413  
QY 2133 VTLAQHIEVITQDTRHHPOOL-SAPLPAPLYSPGASCPVLDLRRPPSDLYLPPPDHG 2191  
Db 2414 --tseelteaktptsspekpqelvtcaeaapstssats-----spep 2455  
QY 2192 APARGSPHSEGRKRPENKTSV-----LGGEDGIEVPSPPEGMTEPGHSRSAYVPLL 2245  
Db 2456 spar-----pprrtsadveirgqgrpg-qppgpkvlrklpgrlvtvv----- 2499  
QY 2246 YRDGQTTPSRMGSGSPGNTSQ-PPAFFSKLTESNSAMVSKK-QEINKKLLTHNNEPE 2303  
Db 2500 -----eekelvrrrrqrgaastlvpgvsetsaspgsvrsmgspesspplgppceaps 2555  
QY 2304 YNISOPGTEIFNMP-----AITGTG-----LMTYRQAQVEH-----ASTN 2339  
Db 2556 ssipppgqgfiairhielgvctggspengdallaitpavkrrrrgrppkknrspadag 2615  
QY 2340 MGLEAIIRKALMGKYDQWESP-PLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPG 2398  
Db 2616 rgvdeapsstlkgktngadpvgpgetlivadpqlipgpplpqpghrpnplisp- 2674  
QY 2399 GGGKAKVSGRPSRRAKS-PAPGLASGDRPPSVSVHSEGCNRRPTLTNRVWEDRPSSA 2457  
Db 2675 --vekrirgrpp--kardlplp-----tissa-gdgnsest-----qppphps 2714  
QY 2458 GSTPPFPYNPLRLQAGV-----MASPP-----PPGLPAGSGP 2490  
Db 2715 pltlp--pllvctatvantvtvtsttsppkrrgrppknppsprp 2760  
  
RESULT 10  
AAB50363  
ID AAB50363 standard; protein; 2972 AA.  
XX  
AC AAB50363;  
XX  
DT 12-MAR-2001 (first entry)  
XX  
DE Human SRCAP.  
XX  
KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;  
KW CAMP regulatory element; CREB binding protein; CBP; ATPase;  
KW transcription activation; DEAD box RNA dependent helicase;  
XX adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.  
OS Homo sapiens.  
XX  
XX WO200073467-A1.  
XX  
XX 07-DEC-2000.  
XX  
XX 25-MAY-2000; 2000WO-US14719.  
XX  
XX 27-MAY-1999; 99US-0136620.  
XX  
XX 25-MAY-2000; 2000US-0579181.  
XX  
XX (UYSL-) UNIV SAINT LOUIS.  
XX  
XX Chrivia J, Yaciuk P;  
XX  
XX WPI; 2001-061545/07.  
XX  
XX N-PSDB; AAC89860.  
XX  
XX Snf2 related cAMP regulatory element (CREB) binding protein (CBP)  
XX activator protein, capable of co-activating CREB binding protein,  
XX useful for modulating transcription and for affecting viral infection -  
PT







```
Qy 1564 ETPR-----LQF-----GSLSSKASQDRKLTSTPREIAK-- 1594
|||
Db 1618 pppspfyldslsleekrkrserlerlqfseahgalavvygtvldfctlpqvaspi 1677
|||
Qy 1595 SPSTVPEH-----HPHTSPYHLLRGVGVLDYRSHIPLAFTDPSISPGIPLD 1644
|||
Db 1678 gprspgshpftwteahravifpqgrldqslsilerfifvmp-----pveapp 1729
|||
Qy 1645 AAAAYLPHLAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677
|||
Db 1730 shachppwlapraaafqeqlasewprarplhrivcnmrtqfplldlydcgkltl 1789
|||
Qy 1678 ALENQ-----TIINDYTSQ-----QMHNNTATAMAQRAQDMLRG----- 1712
|||
Db 1790 avllrqkaeghrvlifqtmrldvleqfityhghlyrlldgstrveqrgalmerfnad 1849
|||
Qy 1713 -----LSPRESSALNAYAGPRGIIDLSQVPHLPVLVPTPGTPATAMDR----- 1757
|||
Db 1850 krifcflitstsgvgvnltagdvtfydsdw-----nptmdaqdqrchrigqtr 1900
|||
Qy 1758 -----LAVLPTA-----PQPFSSRHS 1773
|||
Db 1901 dvhyrlisertveenilkkanqkmlqdmaleggnftayfkqqtirelfdmpleepss 1960
|||
Qy 1774 SS-PLSP-----GGPHTLTKPTTSSSERDR----- 1800
|||
Db 1961 ssvpsapeeeetvaskqthillealcradeedeiraatqakaevaelaefnendgfa 2020
|||
Qy 1801 -----DRDRDRDRREKSILSTTTVEH----- 1823
|||
Db 2021 gegeagrpgadeemsraeqeiaalveqltpieryanmfleasleevsrealkgaeqv 2080
|||
Qy 1824 -----APIWR-PGTQSSSG--SSGGGGSSSRPASHSHAHOH--SPISPR 1867
|||
Db 2081 eaarkldqakeevfvlpqeeegpgagadesctg9gthrrskakapergtvrsrl 2140
|||
Qy 1868 QDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST-----SSVPRPAATFPATHCPLGGT 1924
|||
Db 2141 rgaaetggnhtp-----visahq-----trsttprcparervprprprtbas 2190
|||
Qy 1925 LDGVYPTLMEVLLPKEAPRVARPERPRADTGHAFAPKPPARSGLPEASSKSGSEPRPL 1984
|||
Db 2191 apaaipal---vpvpsavpvlsapnptiilpvhilpspppsqipccspsa--ctpppa 2245
|||
Qy 1985 VPPVSGHATARTPAKNLAPHASP-----DPPAPPASAS-----DPHREKTSQKPSI 2033
|||
Db 2246 ctpppahp---ppaqtclvtspsp1llgppsvpisasvtnlplglrpeaelcaqalasp 2302
|||
Qy 2034 QLELRSLGYHGSS-----YSPEGVEPVs-----SPSLTHDKG-LPKHLE 2075
|||
Db 2303 eslelasvasstetsslsivpkdlipvaveilpvseknsltpsapstleagspngqe 2362
|||
Qy 2076 ELDKSHLGEELRPQPGVKLGGEAAHLP-----HLRPLPESQSPSSPLLQOTAPGVKGH 2129
|||
Db 2363 qeapdsaeqtlitvlp-----egeelplcvseengielppsaasdeplgleadr-- 2413
|||
Qy 2130 QRVVTLAQHISEVLTQDTRHHPOQL--SAPLPAPLYSPFGASCPVLDLRRPDSLILPPP 2188
|||
Db 2414 -----tseelieaktptsspekpgeivcaevaapstssatssp----- 2452
|||
Qy 2189 DHGAPARGSPHSEGGKRSPEPNKTSVLGGGEGDIPVSPPEGWTPEPGHSRAVYPLLXRD 2248
|||
Db 2453 -----egpsarprr-----rrt 2465
|||
Qy 2249 GEQTEPSRMGSKSPGNTSQPPA--FFSKLTESNAMSVKSKOEINKKLTNHNPEYNI 2306
|||
Db 2466 sadveirgggtgrpq--qpgpkvlrklpgrlvtvveekelvqrr----- 2509
|||
Qy 2307 SOPGPEIFNMPALITGLTYRSQAQVQEHASTNMGLEAIRKALMGKYDQWEEPSPLS-- 2364
|||
Db 2510 qqrdaastlvpvgsuet-----saspgspav--rsmsgp-----esspgpigg 2549
|||
Qy 2365 -----ANAFNPLNASASLPAAMP--ITAAGRSDDHTLT 2395
|||
```

```
Db 2550 ceaapssslptppqpfiairrhiehlvgvggspengdgallaitpavkrrgrppkknr 2609
|||
Qy 2396 SPGGGKAKVSGRPSRKAKS-----PAPG-----LASGDRPPSVSSVHSEB 2437
|||
Db 2610 spadagrgvdeapsstlkgktngadvppgtellivadpvlpqplpgqpvh--- 2666
|||
Qy 2438 DCNRRTPLTNRVWE---DRPSSAGSTPPFPYNPLIMRLQAGVMAS-----ppppgLP 2485
|||
Db 2667 ---rppllspekrgrppkardlp-----gtissagdgnsertqppph 2713
|||
Qy 2486 AGSGPLAGPHHAWDEEPKPLLC 2507
|||
Db 2714 spltp1-----ppllv 2725
|||

RESULT 11
AABI2000
ID AABI2000 standard; Protein: 2091 AA.
XX
AC AABI2000;
XX
DT 19-DEC-2000 (first entry)
XX
DE Rat p3103 protein.
XX
KW Rat; p3103 protein; Synonym: SH3 domain; PDZ domain; SAPAP1 interaction;
KW neuronal function; nervous system; neurological disorder.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Domain 554..613
FT Domain /note= "SH3 domain"
FT Domain 654..749
FT Domain /note= "PDZ domain"
XX
XX JP2000184884-A.
XX
PD 04-JUL-2000.
XX
PF 16-NOV-1998; 98JP-0325657.
XX
PR 16-NOV-1998; 98JP-0325657.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2000-567637/53.
XX N-PSDB; AAA62000.
XX
XX A protein Synonym -
XX
XX Claim 1; Page 5-10; 13pp; Japanese.
XX
XX This sequence represents rat p3103 protein, also referred to as
XX Synonym in the title of the specification. The p3103 protein contains
XX both an SH3 domain and a PDZ domain, and was identified as being able to
XX interact with the protein SAPAP1 via the yeast two-hybrid system. P3103
XX protein is therefore thought to be involved in neuronal function, and may
XX be useful for the study of the human nervous system, and for the
XX diagnosis, prevention and treatment of various neurological disorders.
XX
SQ Sequence 2091 AA;
```

```
Query Match 3.2%; Score 429; DB 21; Length 2091;
Best Local Similarity 20.8%; Pred. No. 1.6e-15;
Matches 384; Conservative 140; Mismatches 627; Indels 694; Gaps 84;

Qy 593 EEAI---TPQSAEL--ASMELNESSRWTE--EEMETAKKGLLEHGRNNSAIARMVGSKTIV 646
|||
Db 754 deavhkkasqakrlpppalslrsktsmtseleemvswpkkkieyegqpaavpsmekrtv 813
```



QY 647 SOCKNFYFNYKKRONLEILOHKLKMEKERNARRKKKAPAAAEAEAFPPVVEDEEME 706  
Db 814 yq-----malnkideil-----aaqqtis-----ase 836  
QY 707 ASGVSGNEEMVEEAEALHASNEVPRGCSGPATVNNSSDTESIPSPHTEAAKDTGONG 766  
Db 837 spgpgg-----laslgkhrpkpffa-----tessfdphhrs-----qps 870  
QY 767 PKPPATLGADGPPGPPPPRTS-----RAPIEPTA---SEATGAPTPPPAPPSA 817  
Db 871 ydrpsf-----pppglmlrksgaeddrrpylappamkfrslsvpggsedippppt 925  
QY 818 PPPVVPKEEKEETAAPVEGEQKPPAAEELAVDTGKAEPEVKSECTEAEAEAGPAK 877  
Db 926 spp-----epyy-----stpa----- 937  
QY 878 KDAEAAEATAEALKAKEKGGSGRATTA-----KSGAPQDSDSSATCSADEVDEAE 930  
Db 938 -----pssgrltspgrgpnpsggplpasspsfddgpppdr 978  
QY 931 GGDKNRLLSPPSLTPTGDPANASPOKPLDLKQKRAAIPPIQVTKVHEPP---RE 987  
Db 979 gggrksl-----yhssalppah-----hhpphhhh 1005  
QY 988 DAAPTAKP-----APPAPPPQNLPQESDAPQPCSSPRGKSRGPAPPA 1030  
Db 1006 hppppqhhhhahpphemetgsgdpprllalqpslrgwrggppstsgapsph 1065  
QY 1031 DKEAFAAEAKLPGDDPCWTSGLPPVPPPREVIKASHPADPSAFSYAPP-----GH 1082  
Db 1066 hsssgsgs---ptdaplry---fqlprr-aasaamyvparsggrkplvkqtkvege 1118  
QY 1083 PLPLGLHDTARVLPDPPTISNPPPLISSAKH---PSVLER---QIGATISQMSVOL 1133  
Db 1119 pkgaisassptpalpr-sepppagpeeknsipitliikapstssgrsgsgstee 1177  
QY 1134 HVPYSEH-----AKAPYGVMTGLPLPMDPKKLAPFGVKQEQQLSPRGQAGP 1180  
Db 1178 eptpdgagggsgspapatsvyppsvptpaspgpatldftsqfgeaalvgaark 1237  
QY 1181 PSLGVPTAQAASVLRGALGVPGSGITKIPSTRVPSDSAITRGSTHTHTPADVLK 1240  
Db 1238 eggwgnearrrstflsdagdedgdsqglg-pgg-ppgprlrhskaldeg----- 1286  
QY 1241 GTRITRIIGDPSRLDRGREDSLPRGHVIEGKGHVLISYEGGMSVTOCSKEDGRSSSGP 1300  
Db 1287 -----mfssaepylriesg-----gssgggygaag-----srtgygsgsd 1322  
QY 1301 PHETAAPKTYDMGRVGRATISSASIEGLM-----GRAIPPER----- 1339  
Db 1323 aftstflprr-plvhpltgkaldpasplgalaareralkeseggtpppprppsp 1380  
QY 1340 -----HSPH--HLKEOHIR-----GSITOGIPHSYVEAQED 1369  
Db 1381 dappptllhshpsphsphahepvirllwgdparrelgyraglsgkekaltasppaars 1440  
QY 1370 YLRR-----BAKLLKREGTPPPPPSRDLTEAYKTOALGPKLKPAAHEGLVATVKEAG 1422  
Db 1441 llhrlpttagvgpplllqgl-peptphgvskawrta----- 1478  
QY 1423 RSIHEIPREELRHTPE-LPLAPRLKEGSGITGTPPLKYDVTGASTGSKHVDVRSIGSPG 1481  
Db 1479 -----pee-----perlphvrien---cqarp-----ppagtrgssted-----gpg 1514  
QY 1482 RTFPVPHPLVDWADARALERACYESLSKRCTASSSGSSTARGAPVIVPELGRQSP 1541  
Db 1515 ---vpppsprvl-----ptsptsrgrneenglpllv-----lpppaps 1550  
QY 1542 TYEDHGAFFAGHLPRGSPVTWREPTPRLEQSLSSKASQDRKLTSTPREIAKSPHSVP 1601  
Db 1551 vdvddg-----eflfaeplppplefsnfek-----pespit 1582  
QY 1602 EHHHPHPSYEHLLRGVSGVDLYRSHIPLAFDPTSPRIGIPLDAAAAYYLRHPLAPNTY 1661

Db 1583 pppphlp-----lilil-----dpps-patplpaa----- 1602  
QY 1662 PHLYPPYILIRGYPDAALLENKQTIINDYITSOQMHNTATAMAQRADMLRGLSPRESSLA 1721  
Db 1603 -----pppavaaappt--ldstassitsy-----dseva 1629  
QY 1722 LNYAAGPRGIIDLSQVPHLPVLPVPTPGTPATAMDRLAYLPTAPQFFSSRRHSSSPLSPGG 1781  
Db 1630 tltggaapaagd-----ppapgpapa-----apappapqp-----g 1661  
QY 1782 PTHLTKPPTTTSSERDRDRDREREKSILTSTTTTVEHAPITWRPTEQSSGSSSS 1841  
Db 1662 p---dpppgtdsgieevdsrsshple-----tissastlsl-----saeggnatgva 1709  
QY 1842 GGGGSSSRPASHAHQHSPISPRTDALQORPSVLHNTGMKGIIITAVEFSKPTVL--- 1898  
Db 1710 gggagvas-----gtelltdyavldqgaifgsggtgppvppqlmtps 1752  
QY 1899 ----RSTSSSPVRPAATFPFATHCPLGGLDGVYPTLMEPVLLPKEAPRVARPERPRAD 1954  
Db 1753 klrgalgtsgnlrpg-----psggldrpdvtpt-----sptvs--vtgagt 1791  
QY 1955 TCHAFILAKPPARGSLPASPSKSGSEPRPLVPPVSGHATIAITPAKNLAPHASDPDPAP 2014  
Db 1792 dgllalsacpstagvagpp---vavevevppvpplpa--asslprkllpweegppppp 1846  
QY 2015 --PASASDPHREKQOS-KPFSIQELESLSLGVHSGSSYSPGVEPVPVSSPSLT--HDKG 2069  
Db 1847 plpgplsqpasalatvksaiselssklqfsgss-taggalpwarggsggstshhg 1905  
QY 2070 ---LPKHEELDCKSHLEGELRPQKQPVKLGEEAAHLPHLRPLPESQPSSS----- 2117  
Db 1906 asyiptertsslqrqlsed-----sqstsl-----skpsssfqwpkpp 1945  
QY 2118 -PLLTQAPGVKHQHVVTTLAQHISEVITQDYTRHHPOOLSAPLPAPLYSFFGASCPVLDL 2176  
Db 1946 lpplptgsgvsstaagagatpasaassastr-hlgvfevemppl----- 1992  
QY 2177 RRPSPDLPLPPDHCAPARGSPHSGGKRSPENKTSVLGGEGDIEVPSPPEGWTEPGH 2236  
Db 1993 rrapssllpasdh-----kvspaprpssl-----plp----- 2021  
QY 2237 SRSAVYPLLYRDGEOTEPE-----SRMGSKSPGNTSOPPAFFSKLTE 2277  
Db 2022 -sgplypglf-dirslqpegrearltpslpslchhtpgylgvllea 2064  
  
RESULT 12  
AAR26049  
ID AAR26049 standard; Protein; 1404 AA.  
XX AC AAR26049;  
XX AC AAR26049;  
DT 02-FEB-1993 (first entry)  
XX MSF precursor.  
XX Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
KW stability; proteolytic cleavage; adhesion; alternative splicing.  
XX Synthetic.  
  
Key Location/Qualifiers  
Region 1..26  
Exon\_I  
Region 26..67  
Exon\_II  
Region 67..107  
Exon\_III  
Region 107..157  
Exon\_IV  
Region 157..200











QY 1440 PLAPRLKESGIT-----QGTPKLYDTGAS-----TTGSKK 1470  
 Db 932 taapkmktettteskitattgtvtsttdtqtpfittktittlaptkvtttkt 991  
 QY 1471 HDVRLSGPQRTFPVHPLDMADARALERACYEESLKSRGTASSGGGS-IARGAPVI 1529  
 Db 992 ittteimkpkpeet-----akpkdratnskattppkpkapktstskkpk 1039  
 QY 1530 VPELGKPRQSPLYEDHGAPAGHLPGRSPVTMRE--PTPLQEGSLSSS---KASQDRK 1584  
 Db 1040 mprvrpkpkip-----tprkmtstmpelntpsriaeamlqttrpndpnsk 1086  
 QY 1585 LT-----STPRIAKSPHSTVPEHHHP-----PISPYEHLRLGSGVDLYRS 1626  
 Db 1087 lvevnpksedaggaetpnhllr-phvfmpvtpmdyprvpnqgliin----- 1136  
 QY 1627 HIPLAFDPTSTPRGIPLD 1644  
 Db 1137 --pmlsdetnfcngkpv 1152

## RESULT 14

AAB29773

ID AAB29773 standard; Protein; 1404 AA.

XX AAB29773;

XX 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF; megakaryocyte stimulating factor; tribonectin;  
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
 KW friction coefficient reduction; gene therapy; antiarthritic;  
 KW osteopathic.

XX Homo sapiens.

XX WO200064930-A2.

XX 02-NOV-2000.

XX 24-APR-2000; 2000WO-US10953.

XX 23-APR-1999; 99US-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Jay GD;

XX WPI: 2001-024673/03.

XX N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating  
 XX osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonectin which is a product of  
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
 CC gene. The tribonectin has at least one O-linked oligosaccharide  
 CC lubricating moiety and has a polypeptide sequence comprising 1-76  
 CC repeats of a motif having at least 50% identity to the sequence KEPAPT  
 CC (AAB29774). The invention also relates to a nucleic acid encoding a  
 CC human MSF-derived tribonectin; a biocompatible composition comprising a  
 CC human tribonectin for inhibiting tissue adhesion formation; and a method  
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
 CC measuring the amount of MSF or its fragment in a biological sample of a  
 CC mammal, wherein an increased amount of MSF compared to a control  
 CC indicates the presence of or predisposition to developing  
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the

CC treatment of osteoarthritis, where they may be used for lubricating  
 CC mammalian joints, such as articulating joints of humans, dogs or horses.  
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
 CC useful for inhibiting adhesion between two surfaces such as the injured  
 CC tissues of a mammal, where the injury is caused by a surgical insertion  
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
 CC tribonectin may be used in gene therapy. The present sequence represents  
 CC human MSF.

XX Sequence 1404 AA;

Query Match 3.2%; Score 428; DB 22; Length 1404;

Best Local Similarity 20.8%; Pred. No. 1e-15;

Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;

QY 476 NENYKSLVRRYRRGKQ-----QQQQQQQQQQQQQQQMPRSSQEEKDEKE 524  
 Db 126 sqtiksttkrspkppnkkkkvieseeiteehsvsenqssssssssstlwkikss 185

QY 525 KEKEAEKE-EKPEYENDKEDLLKEK-----TDDT-SGEDNDEKAVASKRKTANSQ 575  
 Db 186 knsaanrelqklkvdknknrtkkktpkppvdeagsgldngdfkvttdtsttqhknk 245

QY 576 GRRKGRITRSMANEANSEEAITP-QQSAELASGMELESRSRWTEEMETAKKGLLEHGRNW 634  
 Db 246 vstspkit--takpinrpslppnsdtsktsltvnkettvetkttntktsdgtgek 303

QY 635 SAIAWMVGKTVSQCKNFYFNKKRONLDEILQOHLKME---KERNARRKKKAPAAAS 691  
 Db 304 ttsaketqsiektakdl-----aptsklaktpkaettkgpaltpkepttpk 356

QY 692 BEAAPPVVEDEMEASGVSGNEEWEAEALHAGSNEVPRGECGATVNNSSDTSI 751  
 Db 357 epasttp-----kept-----pttiksaapttkpe 380

QY 752 PSPHTFAAKDTGQNGPKPPATLGADGPPPPPTPRRTSRAPIEPTPASEATGAPTPP-- 809  
 Db 381 paptttksaapttkpeapttt-----kepapttkpeaptttkpeaptt-tksaapttkpe 434

QY 810 PAPPSPPSAPPVVPVKEKEEETAAAPPVEEGEEKPPAAEEAELAVDTGKAEVPKSECTEE 869  
 Db 435 paptttkpkpapttkpe--papttkpepttkpeaptttkpeaptttkpeaptt----- 485

QY 870 AEEGPAKGDAAEAATAEGALKAEKKGSGRATKSSGAPQSDSSATCSADEVDEA 929  
 Db 486 apkpkpapttkpeaptttkpeaptttkpe-----pspttkpeaptttksaapttkpe----- 537

QY 930 EGGDKNRLISPRPSLLTPTGDPPRANASPKPLDLKQLKQRAAAIPPIQVTKVHEP----- 985  
 Db 538 -----aptttksaapttkpe-----aptttkpeaptttkpeaptttkpeaptttkpe 565

QY 986 REDA--APTTPAPAPPQPNQLQPSDAPQOQPGSSPRKSRSPAPADKEAFAPAEAKLIP 1043  
 Db 566 kepapttkpkpapttkpeaptttkpkpapttkkpkpapttkkpkpapttkpkpapttkpkpklit 625

QY 1044 GDPP---CWTSGLPFPVPPREVIVKASPHADPPSAFSAFYPAGHPPLPLGLHDFARVLP 1099  
 Db 626 pttpeklapttkpeaptttkpeelapttkpepttkp-----peepapt-tpkaaapntpkpe 679

QY 1100 PTISNP-PPLISSAKHPVLERIGAIQSGMSVQLHVPYSEHAKAP-----VGPVTMGLP 1153  
 Db 680 paptttkpeaptttkpea-----pttpketapttkpgtapttkp 720

QY 1154 LPMDPKKLAPSGVKGQQLSPRGQAGPPESLG---VPTAQEASVLRGTALGALGVPGSGITK 1210  
 Db 721 apttkpkp-----kelapttktsdskpapttkp-----kgta----- 758

QY 1211 GIPSTRVPSDAITVRSITHTGTPADVLVKGITVILIGEDSPSLDGRGDSLPKGVIIY 1270  
 Db 759 --ptt--pkpapt-----tpkeaptttkgtapttkpkeaptttkpkpkelap----- 804







Db 1097 rrsqssteaepdqdgagggsspsapatsvppspvptpaspsgpatlftsf 1156  
QY 1171 QLSRQAGPPESLGVPTAQEASVLRGTALGSPVGGSIKGPSTRVPSDSAITVRGSI 1230  
Db 1157 gaalvaarregwqearrrtflstfdagdedgdsqg-p9g--ppgprlrhsksid 1213  
QY 1231 HGTADVLKGYITRIIGEDSPSRDLDRGDSLPKGVHYYEKKGHVLSYEGGMSVTQCS 1290  
Db 1214 eg-----mfaeapylriesg-----gssggygayaag-----s 1241  
QY 1291 KEDGRSSGPPHETAAPKRTYDMEGRGRAISSASIEGLM-----GRAI 1335  
Db 1242 rayggsgsssaftflppr--plvhlqkalpasplglalaareralkesseggtq 1299  
QY 1336 PPER-----HSPH--HLKQKHIR-----GSITQCI 1359  
Db 1300 ppprpsprydappptllhshpsphsharhepvrlrhwgdparrtelgyraglsgkeal 1359  
QY 1360 PRSYVEAQEDYLRR-----EAKLLKREGTTPPPPSRDLTEAYKTQALGPLKPAHE 1412  
Db 1360 tasppaarrrllhrpptagvgvplllqig-pepptphpgvskawrtaa----- 1407  
QY 1413 GUVATWKEAGRSIHETPRELRHTE-LPLAPRLKEGSITQGTPLKYDTGASTTGSKKH 1471  
Db 1408 -----pee-----periplhvrflen---cqp-----ppagtrgsste 1439  
QY 1472 DVRSLTGSPGRTFPVPHPLDVNMADARALERACYEESLKRPGCTASSGGSIARGAPVIVP 1531  
Db 1440 d-----gpg--vppsprvl-----ptspsrpgneenglpllv- 1472  
QY 1532 ELGKRQSPLTVEDGAPFAGHLPRGSPVTMRPRTLRQESLSKSKASQDRKLJSTPRE 1591  
Db 1473 ---lppapsvdvddg-----eflfaepipplefnsfek----- 1505  
QY 1592 IAKSPHSTVPEHHHPHPSYEHLLRCVGVLDLYRSHIPLAFDPTSIPIRGIPLDAAAAYYL 1651  
Db 1506 ----pesplitpgpphlp-----dpps--patpipa----- 1531  
QY 1652 PRHLAPNTYPHYLYIRGYPTAALENROTIINDYITSOOMHNTATAMAQRADMLR 1711  
Db 1532 -----ppavaaaapt--ldstassitsy----- 1553  
QY 1712 GLSPRESSLALNYAGPRGIIDLSQVPHLPVLVPTPGTPATAMDRLAYLPTAPQPFSSR 1771  
Db 1554 -----dsevatltqgapaapg-----ppapgpapa---apabpapg----- 1589  
QY 1772 HSSSLPSGPGTHLTKPTTSSSERERDRDRDREREKSIILTSTTVEHAPIWRPCT 1831  
Db 1590 -----gp---dppp9tdsgieevdsrssdhple-----tissastlssl-----sa 1628  
QY 1832 EQSGSGSGSGSGSGSGSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKIITAVE 1891  
Db 1629 eggngtgvaggagvan-----gtelldtyayldqqaqfgsggtppg 1671  
QY 1892 PSKPTVL-----RSTSTSSVPRPAATFPATHCPGLGTLGVYPTLMPEVLLPKEAPR 1944  
Db 1672 pyppqlmtpsklrgalgtsgnlrpg-----psggldrqpvtpt-----spt 1712  
QY 1945 VARPERPRADTGHAFLAKPPARSGLEPASSPSKSEPRPLVPVSGHATIARTPAKNLAP 2004  
Db 1713 vs--vtgagtdgllalsacpgpstagvaggp---vavepevpvpplpa--asslprkilp 1765  
QY 2005 HHASDPDPAP--PASASDPHREKTS-QPFSIQELELSLGVHSGSSYEGVEVPSPVSS 2061  
Db 1766 weegppppppplpgplsqasalatvkaasielssklqqfggss--taggalpwarggs 1824  
QY 2062 PSLT--HDKG-----LPKHLELDKSHLEGELRPKPCPVKLGGEAAHPLHRLPLPESQFSS 2116  
Db 1825 ggstdshggasyipertsslqrqlrsed-----sqtssl-----skpss 1864  
QY 2117 S-----PLLOTAPGVKGHQRVVTLAQHISEVITQDYTRHHHPQQLSAPLPAPLYSF 2166

Db 1865 sifqnwpkplpplptgsgvsstaapatspsassasastr-hlqgvfemrplpl---- 1920  
QY 2167 PGASCPVLDLRRPPSDLYLPPDPDHGAPARGSPHSEGGKSPENKTSV----- 2214  
Db 1921 -----lrrapssllpasdh-----kvspaprsslplpsgpiypgl 1958  
QY 2215 -----LGGEDGIEPV-SPPE-----GMTEPGHSRSVAVYPLLYRDGEOTEPSR 2256  
Db 1959 fdirsstptgaggstdpfapvfvpphpglsgglggalsgasrs-----lsptr 2006  
QY 2257 MGSKSPGNT--SOPPAFFSK 2274  
Db 2007 iislpdpkpfgakplgfwtk 2026  
  
Search completed: September 8, 2001, 14:34:13  
Job time: 17730 sec



---







Db 70 PSGKAQPHASVYVSEAGKDKGPPKSRVEELRTRGKTTITAANFIDVTITRQIASDKDA 129  
QY 2022 HREKTSQKPFSTOELELSLGYHSGSYSPGVEPVSPVSSPLTHDKGLPKHLELDKSH 2081  
Db 130 RERGSSDSS-----SSLSHRYETASDALEVISPASSAPPOEKPOAYOPDMVKANQ 183  
QY 2082 LEGERLPKQPGVKLGGEAAHPLHLR-----LHPESQPSSSPLLOT--APGVKGHQRVVT 2134  
Db 184 AENESTROYEGP-----LHHYRSQOESPSPQOQPLPSSQSEGGMQVPRTHRLIT 234  
QY 2135 LAQHSITVITQDTRHHPOOLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189  
Db 235 LADHICQIITODFARN--QVPSQSTSTFTQTSFALSSTPV---RTKTSRYSSESQST 289  
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGGEGIEPVSPPEGMTEPG-HSRS 2240  
Db 290 VUHPRGPRVSPENLVDRSGRSPKSPERSHI---PSEPIEPIPPQG---PAYHEKOD 343  
QY 2241 VYPLLYRDGEQTEPRMGSKSPGNTSQPAFFSKLTESAMVSKKQOINKKLTHNRN 2300  
Db 344 SMLLSQRGVDPAEQRSRSPGSIYLPSPFTKL-ESTSPMVSKKQOIFRKLNSGGG 402  
QY 2301 EPEYNISQPGTEIFNMPAITGTLTYRSQAOVEHASTNMGLEAIRKALMGKYDQWEE- 2359  
Db 403 DSDMAAQPGTEIFNLPAVTTSGAVSSRSHPADPAS-NLGLIEDIRKALMGFDDKVED 461  
QY 2360 -----SPPLSANAFNPLNASILPAAMPITTAADGRSDHLLTSPGGG-GKAKVSGRPSRK 2413  
Db 462 HGVVMSHPV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 513  
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRTPLTNRVWEDRPPSAGSTPPFPYNPLIMRL 2471  
Db 514 SKSPIPGQSYLGTERTPSSVSVHSEGDYHROTP--GWAWEDRPSSTGSTQFPYNPLTIRM 571  
QY 2472 QAGVNASPPPPGLPAGSGPL--AGPH---HANDEPKPLLCQYETLSDSE 2517  
Db 572 -----LSSTPTQIACAPSAITOAAPHQONRIWEREPAPLLSAQYETLSDSD 618

## RESULT 2

PCT-US95-16311-5  
; Sequence 5, Application PC/TUS9516311  
; GENERAL INFORMATION:  
; APPLICANT: Moore, David  
; APPLICANT: Seol, Wongi  
; APPLICANT: Choi, Hueng-Sik  
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street, Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16311  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 619 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
PCT-US95-16311-5

## Query Match 5.4%; Score 718; DB 5; Length 619;

Best Local Similarity 36.0%; Pred. No. 1,5e-33;  
Matches 213; Conservative 76; Mismatches 212; Indels 90; Gaps 26;

QY 1975 PSKGSEPRPLV-----PPVSGHATIARTPAKN--LAPHHASDPDPAPPASADP 2021  
Db 70 PSGKAQPHASVYVSEAGKDKGPPKSRVEELRTRGKTTITAANFIDVTITRQIASDKDA 129  
QY 2022 HREKTSQKPFSTOELELSLGYHSGSYSPGVEPVSPVSSPLTHDKGLPKHLELDKSH 2081  
Db 130 RERGSSDSS-----SSLSHRYETASDALEVISPASSAPPOEKPOAYOPDMVKANQ 183  
QY 2082 LEGERLPKQPGVKLGGEAAHPLHLR-----LHPESQPSSSPLLOT--APGVKGHQRVVT 2134  
Db 184 AENESTROYEGP-----LHHYRSQOESPSPQOQPLPSSQSEGGMQVPRTHRLIT 234  
QY 2135 LAQHSITVITQDTRHHPOOLSAPLAPLYSPGA--SCPVLDLRRPPSDLYLPPD--- 2189  
Db 235 LADHICQIITODFARN--QVPSQSTSTFTQTSFALSSTPV---RTKTSRYSSESQST 289  
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGGEGIEPVSPPEGMTEPG-HSRS 2240  
Db 290 VUHPRGPRVSPENLVDRSGRSPKSPERSHI---PSEPIEPIPPQG---PAYHEKOD 343  
QY 2241 VYPLLYRDGEQTEPRMGSKSPGNTSQPAFFSKLTESAMVSKKQOINKKLTHNRN 2300  
Db 344 SMLLSQRGVDPAEQRSRSPGSIYLPSPFTKL-ESTSPMVSKKQOIFRKLNSGGG 402  
QY 2301 EPEYNISQPGTEIFNMPAITGTLTYRSQAOVEHASTNMGLEAIRKALMGKYDQWEE- 2359  
Db 403 DSDMAAQPGTEIFNLPAVTTSGAVSSRSHPADPAS-NLGLIEDIRKALMGFDDKVED 461  
QY 2360 -----SPPLSANAFNPLNASILPAAMPITTAADGRSDHLLTSPGGG-GKAKVSGRPSRK 2413  
Db 462 HGVVMSHPV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 513  
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRTPLTNRVWEDRPPSAGSTPPFPYNPLIMRL 2471  
Db 514 SKSPIPGQSYLGTERTPSSVSVHSEGDYHROTP--GWAWEDRPSSTGSTQFPYNPLTIRM 571  
QY 2472 QAGVNASPPPPGLPAGSGPL--AGPH---HANDEPKPLLCQYETLSDSE 2517  
Db 572 -----LSSTPTQIACAPSAITOAAPHQONRIWEREPAPLLSAQYETLSDSD 618

## RESULT 3

US-08-061-376-5  
; Sequence 5, Application US/08061376  
; Patent No. 6175000  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; APPLICANT: Djabali, Malek  
; APPLICANT: Selleri, Lucia  
; APPLICANT: Parry, Pauline  
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23  
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles



STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/061.376  
FILING DATE: 13-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9387  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)546-4737  
TELEFAX: (619)546-9392  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3969 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-061-376-5

Query Match 2.94; Score 380; DB 4; Length 3969;

Best Local Similarity 18.09; Pred. No. 3.3e-13;

Matches 590; Conservative 357; Mismatches 1082; Indels 1250; Gaps 148;

QY 51 SHLSPGSIHQORRPSLLS-EFQGNERSQELHLRPHSHVLP-----ELGSEMEFTES 105  
DB 679 SPLSGTFRDMHKRPLLRPRFT-----SEAHSRIFESVTLPSNRTSAGTSSSGVSNR 733  
QY 106 KRPLELLPDLRLPSPLLATGQAGSEDLTKDRSLTKGLEPVPSPPHPTDPELELYPP 165  
DB 734 KKKRVFSP---IRSEP---RSPSHSMRTSRGLSSLSSELSPLTPSPSSLSLSVSP- 785  
QY 166 RLSKEELLQNDVRDREITVMEVQIISKKKQOQLEEAAPPEPEKPVSPPTTESKURS 225  
DB 786 -LATSALAPTFPTFPHSHTL-----QSGSAEKQNRPKOTSAPAEFPSSSPTPL----- 834  
QY 226 LVQIIYDNRKKAHAHRLLEGGLGQVELPLYNQPSDTRQYHENIKINQAMRKILLYFK 285  
DB 835 -----FPWFTPGSQTERGNKDKAPEELSKD----- 860  
QY 286 RRNHARKQWKQFCQRYDQLEALEKKVERIENPNRRRAKESKVREYEEKQFPEIRKQRE 345  
DB 861 -ROADKSVEKDK-----SRERDREREKENRESKEKR-----KKGEISQSSA 903  
QY 346 L-----QERMQRVQRGSGLSMSAAR-SEHEVSEIIDG-----LS 380  
DB 904 LYPVGRVSKEVVGEDVATSSSAKATGRKKSSSHSDGTDITSVTLGDTTAVTKILIKK 963  
QY 381 EOENLEKQMRQVLAVIPPMYLDADQORIKFINNGLMAOPMKVYKDRQVMNMWSEQKETF 440  
DB 964 GRNLEKNLNDLGPATPSL-----EKEKTLCLTSPSSSTVK-HSTSSIGSLMAQADKLPM 1017  
QY 441 REKFMQHPKFNGLIASFLERKTVAECVLYIYITKKNENYKTVLRRSYRRRGRKSGQQQQQQ 500  
DB 1018 TDK-----RVASLLKKAQALC-----KIEKSKSL----- 1042  
QY 501 QQQQQQQQQQWPRSSQEKEDEK-----EKEAEKEEKPEVE-- 539  
DB 1043 -----KQTDQPAQAGQSSDSSETSVRGPRIKHYCRRAAVALGRKRAVFPDDMPTLSAL 1095  
QY 540 --NDKEDLLKEKTDGTDSCEDNDEKAVASKRKTANSQGRKKGRITRSMANFANSEAI- 596  
DB 1096 PWEEKIL-----SSMGNDKSSIA--GSEDAEPLAPPIKIPVTRNKAPQEPVVK 1146

QY 597 ---TPQSAELASMELNESSRWTEEMETAKKGLLEHGRNWSAIAIMVSGSKTVSOCKNFY 653  
DB 1147 KGRRSRCGQCPGQVPEDCGVCTNCLDKPKFG-----GRN-----IKQC----- 1187  
QY 654 FNYKQRNLDEILQOHLKMEKERNARRKKKAPAAASEAAFPVVEDEMEASGVSN.713  
DB 1188 CKMRKQNLQ--WMPSKAYLQAKAVKKKKKTS-----EKKDSKSSVVK 1235  
QY 714 EEMVEEAALHASGNEVPRGCSG-----PATVNNSSDTEIPSPH-----TEAAKD 761  
DB 1236 ---VYDSQKPTPSAREDPAPKSSSEPPPRKPVEEKSEGNVSAPGPEKSKATTPASRK 1292  
QY 762 TGQNGPKPPATLGADGPPPGPT--PPRRTSRAPTEPASEATGAPTPPPAPPSPS--- 816  
DB 1293 SSKQVSQPALVI-----PPQPTTGPRK--EVP-KTTPSEPKKKOPPPPESPQSKOK 1344  
QY 817 --APPVVPKEKEEETAAAPVEGE-----EOKPPA----- 847  
DB 1345 KVAPRPSIPVKQPKPEKEKPPVNVKQENAGTNILSTLSNGNSSKOKIPADGVHRIRVDF 1404  
QY 848 ---AEE-----LAVDTGKAE-----EPVKSECTEEAE 872  
DB 1405 KEDCEAENWEMGGIGILTSPVITPRVVCFCASSGHVEFVYVCVCEPFHFHCFLEE-NE 1463  
QY 873 GPK -----GKDAEAAEATAE-----GALKAEKKE-- 897  
DB 1464 RPLEDLENWCCRCKFCHVCGRQHOATKQLECNKCRNSYHPECLGPNYPTRPKKKKV 1523  
QY 898 -----GSGRATTAKSSGAPQDSSSA-----TCSADEVEAE 930  
DB 1524 WICTKVRCKSCGSTTPGKWDQASHDFSLCHDCAKLFAGNFCPLCDKCYDDDDYDSEK 1583  
QY 931 -----GGDKNRLLSPPRLTTP-----GDPRANASPKQLDL 963  
DB 1584 MMQCKDCDRWHKSCENLSDEMYEILSNLPESVATVCTNCTERHFAEMRLALEKELQISL 1643  
QY 964 KQ-----LKORAAAIPIQVTKVHEPRDEAATKPAAPPQPQNLQPS 1009  
DB 1644 KQVLTALLNSRTTSHLLRYQAAPDDL-----NPETEESIFSRSSPEGPDPPVLTEVSK 1698  
QY 1010 DAPQPGSSPRGKSRPAPPADKEAFABAQKLPDGPWCWTSGLFPVPPPRVVIKASPIA 1069  
DB 1699 QDQOQP-----LDLEGVKKRMDQGNITSVLFEFSDDDIVKIIQAANS 1739  
QY 1070 PPSAFSVAPGHPPLPLGLHDTARP-----VLP-----RPPTISNPPPLISS 1111  
DB 1740 DG-----GQPIKKANSVMKSFFFIOMERVEFWFSVKSRFWEPNKNVSSNGMLPN 1790  
QY 1112 AKHPSVL-----ERQIGA-ISQMSYQLHVYPYSEHAKAPVGPVTMGLPLPMDPKKLA 1162  
DB 1791 AVLPPSLDHNVAQWQEREENSHTEQPLMKKIIP---APKPGGEPDPTPLHPPTPP 1846  
QY 1163 PFGVKQQLSPRQAGGPESL-----GVPTAQEA----- 1192  
DB 1847 ILSTDREDSRP--ELNPPPGIEDNRQCALCLTYGDDSDANDAGRLLYIGQNEWHVNCAL 1904  
QY 1193 -----SVLRGTALG---SVPGGSITKIPSTRVPSDSAITYRGST 1230  
DB 1905 WSAEVEFDDGSLKNVHVAIRGKQLRCEFCOKPATV--GCCLTSCTSN----- 1952  
QY 1231 HGTPADVLYKQITR-----IIGEDSPSLDRGREDLSLPGHVI-----YEG 1272  
DB 1953 -----YHFMCSRAKNCVFLDDKKVYQHRD--LIKGEVVPENGFEVFRVFDPEG 2002  
QY 1273 -----KGHVLSYEGMSVT-----QCSKEDGRSSSGPP 1301  
DB 2003 ISLRKFLNGLEPENIHMWIGSMITDCLGILNDLSDCEDKLFPIGYQCS----- 2051  
QY 1302 HETAAPKPTYDMMEGRVRAISSASIEGLMGRAIPERHSPHHLKEQHHRISITQIGIPR 1361  
DB 2052 -----RVYSTTTPDARKCVYTKIIVECRPPVPPEPDINSTVEHDENTIAHS-----PT 2099  
QY 1362 SYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAY-----KTOALGPLKLPKPAH 1411







Db 625 GPP-----GAPG-----PAGPPGSRGDPGPPGAPG-----PAGPP 654

```

; CURRENT APPLICATION DATA:
.
```







;; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
;; FILE OF INVENTION: PREPARATION THEREOF  
;; FILE REFERENCE: 2728-2  
;; CURRENT APPLICATION NUMBER: US/09/219,849  
;; CURRENT FILING DATE: 1998-12-23  
;; NUMBER OF SEQ ID NOS: 50  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 5  
;; LENGTH: 960  
;; TYPE: DPT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Illustrative  
;; OTHER INFORMATION: amino acid sequence  
US-09-219-849-5

Query Match 2.8%; Score 365; DB 4; Length 960;  
Best Local Similarity 21.3%; Pred. No. 4e-13;  
Matches 337; Conservative 33; Mismatches 490; Indels 720; Gaps 78;  
QY 726 ASGNEVPGECSGATVNSSD--TESIPSPHTEAAKTGQNGKPPATLGAADGP--- 779  
DB 11 AHGPAGPKG-APGPAGPGSGRDPGPPGAPG-AGPPGSRDPGPPGAPGPPGSGRD 65  
QY 780 PGPP-----TPPRTSRAPTEPTPASE-----ATGAPT--PPA---PPSPSAPPP 820  
DB 66 PGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDPGPPGAPG 125  
QY 821 VVPKEEKEETAAPVVEEGEQKPAEELAVDTGKAEVVKSECTEAEAGPAKGD 880  
DB 126 AGPPGSRDPGPPGAP-PPGSRDPGPPGAPGPPGSRDPGPPGAPGPPGSRDPGPP- 168  
QY 881 EAAETAEALKAERKEGSGRATTAKSSGAPQSDSSATCSADEVDEAGDKNRLSP 940  
DB 169 GPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDPGPP- 221  
QY 941 RPSLTPTGDPANASPOKLDLKLKORAAAIPTQVTKVHEPPREDAAP-----T 992  
DB 222 -PGAHPAG- PKGAHPAGP-----KGAHPAGPKGAHPAGPKGAHPAGPPGSR 270  
QY 993 KPAPPAPPPQNLQESDAPQPGS-SPRGK--SRSPAPPADKEAFAEAKLPDGPQW 1049  
DB 271 DPGGAPGAPGAP-PPGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDPGPP- 326  
QY 1050 TSGLEPPVPPREVIRKASPHADPSAFSAPPCHPLPLGLHDTARVLPPTISNPPPLI 1109  
DB 327 --GAPGAPGAP-----PGSRDP-----GPPGAPGAPGPPGSRDPGPPGAPGAPGPP- 369  
QY 1110 SSAKHPVLERIGAISSQMSVQLHPVYSEHAKAPVPTMGLPLPMDPKLAPFSGVKQ 1169  
DB 370 -----GSRDPG-----PPGAPGAPG-----PGSRDP----- 391  
QY 1170 EQLSRGAGPPESLGVTQAQASVLRGALSGVPGSGITKIPSTRVPSDAIYRGSI 1229  
DB 392 -----GPPGAPGAPGPPGSRDP-----GPPGAPGAPGPPGSRDPGPP- 428  
QY 1230 THGTADVLTKTITRIGEDSPSLDRCDREDSLPKGVHVIYEGKGVLSYEGGMSVTQC 1289  
DB 429 --GAP-----GPAGPP-----GSRDPGPPG-----APGAPG----- 452  
QY 1290 SKEDGRSSSGPP--HETAAPKRTYDMMEGRVRAISSASIEGLMGRAPIPERHSPHLLKE 1347  
DB 453 --PPSRDPGPPGAPGAPGAPK-----CAHPGAPGKGAHPGAPGAPGK 490  
QY 1348 QHHIRGSITQIGIPRSYVEAQEDYLRREAKLLKREGTPPP--PPPSRDTEAYKTOALGPL 1405  
DB 491 AHGPAG-----KGAPGAPGPPGSRD----- 511  
QY 1406 KUKPAHEGLVATVKEAGRSIHPIPRELHRTPELPLAPRLKEGSTITQCTPLKYDTGAST 1465  
DB 512 -----PGPPGAPG-----AGPPGSRDPGPP- 532

QY 1466 TGSKKHVRSLIGSPGRTFPVPHLDVMADARALERACYEESLKRPGTASSSGGSIARG 1525  
DB 533 -----GAPGAPGPP-----GSRDPG 547  
QY 1526 APVIVPELKGKQSPPLTYEDHGAPFAGHLPGRGSPVYTMREPTRLQEGSLSSKASODRKL 1585  
DB 548 PPGAPGAPGPPGSG-----RDPGPPGAPG-PAAGPPGSG-RDPGPPGAPG-AGPPGSRD--- 596  
QY 1586 TSTPREIAKSPHSTVPEHHHPHPISEYHLLRGVSGVDLYRSHPLAFDPTSPRGIPDLA 1645  
DB 597 -----PGPPGAPGAPG-----PGSRDPG-PPG----- 618  
QY 1646 AAAYLPRHLAPNPTYPHLYPPYLINGYDPTAALENROTIIINDYITTSQQMHHTATAMAQ 1705  
DB 619 -----APGP----- 622  
QY 1706 RADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPTPTGTPATA---MDRLAYLP 1762  
DB 623 -----AGPPGSRD-----PGPPGAPGAPGPPGSRDPGPP 651  
QY 1763 TAPQPFSSRHSSPLSPGPGTHLTKTPTTSSSERERDRDRDREREKSIILTSTTIVE 1822  
DB 652 GAPGAPGPPGSRDPGPPGAP----- 671  
QY 1823 HAP1WRPGTEQSSGSSGSGGGGSSSR---PASHSHA---HQHSPISPTQDALQORP 1875  
DB 672 -GPAGPPGSG-RDPGPPGAPGAPGPPGSRDPGPPGAPGAPGAPGAPGPK----- 720  
QY 1876 SVLHNTGMKGIIITAVEPSKPTVLRSTSSVVRPAATFPFATHCPGLGGLDGVYPTLMEP 1935  
DB 721 -----GAHG-----PAG--PKGAHPAG----- 736  
QY 1936 VLLPKEAPRVARPERPRADTGHAFKAPKPARSGLEPASSPSKGE-----PRPLVPPV 1988  
DB 737 ---PKAGPAGPAGPSR-DPG-----PPGAPG--PAGPP--GSRDPGPPGAPGAPGPPG 782  
QY 1989 SGHATARTPAKNLAPHHASP-----DP-----PAPPASADPHREKTQSKPFSIQELE 2037  
DB 783 S-----RDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPGSRDP----- 820  
QY 2038 LRSGLYHGSYSPEGVE-PVSPVSPSLTHDKGLPKHLEELDKSHLEGLRKPQGPVKL 2096  
DB 821 -----GPPGAPGAPGPPGSG-----RDPG-----PFGAPGAPG 848  
QY 2097 GGEAAHLPLRPESQPS-SPLLOTAPGVKGHQRVVTLAQHSIVITQDTRHHHPQOL 2155  
DB 849 PGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGAPG-----PPGS 885  
QY 2156 SAPLPAPLYSFFGASCPVLDLRRPDSLYLPPPDHGAPARGSPHSEGGKRSPEPNKTSVL 2215  
DB 886 RDPGP-----PGAPGPA---GPPGSRDPGPP--CAPGAPG---PGSRDPGPP----- 924  
QY 2216 GGGEGIEVPSPPEGMTEPG 2235  
DB 925 -PGAPG--PAGPP-GSRDPG 940

## RESULT 7

US-07-853-913-4  
; Sequence 4, Application US/07853913  
; Patent No. 533839  
; GENERAL INFORMATION:  
; APPLICANT: McKay, Ronald D.G.  
; APPLICANT: Lendahl, Urban  
; TITLE OF INVENTION: Nestin Expression As An Indicator of  
; TITLE OF INVENTION: Neuroepithelial Tumors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.



```

: ZIP: 02173
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/853,913
: FILING DATE: 19920319
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/660,412
: FILING DATE: 22-FEB-1991
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/603,803
: FILING DATE: 25-OCT-1990
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/201,762
: FILING DATE: 02-JUN-1988
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/180,548
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-4641AAAA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1618 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-07-853-913-4

```

```

Query Match      2.7%; Score 358; DB 1; Length 1618;
Best Local Similarity 21.1%; Pred. No. 1.9e-13;
Matches 323; Conservative 196; Mismatches 511; Indels 502; Gaps 77;

QY 97 KSEMEFIESKRPRL--ELLP--DPLLPSPL--LATGQPA--GSEDLTKDRSLTGKLEP 147
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 329 KLELOFRTPEGRIIGSLLPVLSPTSPLPATLETPVPAFLKNQEPLOARTPTLASTP 388
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 148 VSPSPPHDTPELELVPPLRSKEELIQMDVRDRHNTVEQOISKL-----KKQOOOLE 201
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 389 I-PPTQAPSP-----AVDAEIRAODAPLSLLOQTQGGKQKOAPEPLR 428
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 202 EEA--AKP---PEPEKP-----VSP-----PPIESKH----- 223
    :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 429 AEARVAIPASVLPGEPEGGQREQEASTQSPEDHIASLAPPLSPDHSSLEAKDGESGRV 488
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 224 -----RSLVQIYDE-----NRKKAERAAH-----RIEGL 248
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 489 FSIERGEQGIWGLVEKETAEIGKVVSQAQEIWEEDLNKREKIQDSQVPLEKETLSL 548
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 249 GPQVE---LPYLNQSPDTRQYHENIKINQAMRKLLILYFKRRNHARKQWKQFCQRYDOL 305
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 549 GEETQESLKTLENOSHETLE--RENQECPRSEEDL-----ET 584
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 306 MEALEKKVERI-----ENPNRRAKESKVREYVEKQFPEIRKORELQERNQSRVGQRG 358
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 585 LKSLEKENKRAIKGGGSETSKRGCRQLKPTGKEDTQTLO--SLQKENQELMKLSLEGNLE 643
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 359 SGLSMSAARSEHEYSEIIDGLSEQENLEKMROLAVTPPMLYDADQOQRIKFINMGLMAD 418
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 644 TFL-FPGTENQELVSSLOENLESITALEKENQEPRLRSEPV--GDEEALRLPTKEN--QE 697
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
QY 419 PMKYVKDRQVNMVMSBQKETFR--EKPQHPKNFGLIASFLERKTVAECVLYYLTKKN 476
    :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 698 PLRSLIED-----ENKEAFRLSEKENQSP-----LKTLEEDQOSTVRPLET 737
    :|:|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|

```

[illegible]



## RESULT 8

US-08-963-825-18  
; Sequence 18, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963.825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187.319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adga C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (I)  
US-08-963-825-18

Query Match 2.7%; Score 351; DB 3; Length 1341;  
Best Local Similarity 20.1%; Pred. No. 3.8e-12;  
Matches 350; Conservative 106; Mismatches 573; Indels 714; Gaps 81;  
QY 722 EALHAGNEVREGCSGATVNNSSDTEIPSPHTAAK-DFGQNGKPPA-TLGADGPP 779  
DB 77 ETKNCPGAEVPEGCC-PVCPDGSSEPTDQETTVGEGRGDTGPRGPGAPGPRGDGIP 135  
QY 780 -----PGPPTPPRRTSRAPTEPTASEATGAPTP-----PA 811  
DB 136 GQGLPGPPGP-----GPPGPGLGNGFAPQLYGYDEKSTGISVPG 178  
QY 812 PPSAP-----PPVVKKEEETAAAPVEEG-----EOKPPAAELAVDTGKA 858  
DB 179 PMGSPGRGLPGPGAPGPGZF-----GPPGZPGZPGASGPMGPRGPPGKBGBGZA 234  
QY 859 EPPVKSCTEEAECPAKDAEAAEAAGALKAEKEGGSGRATTAKSSCAPODSS 918  
DB 235 GKPR-----PGZRPFGPGZGARGLPGTA--GLPMKGHRGFGSLBAGKAGBAGPAG---- 283  
QY 919 ATCSADEVEAEGGDKNLLSPRPSLLPTGDPANASFOKPKPLDLKQLKRAAAIPIQV 978

DB 284 -----PKGZPGS----- 290  
QY 979 TKVHEPPREDAAPTKPAAPPQPNQLQESDAPQPGSPRCKSRSPAPPADKEAFAAE 1038  
DB 291 -----PGZBAGPGZMGP-----GPGNSGEPGAGSK----- 318  
QY 1039 AOKLPDPPCWTSGLPFPFPPPREVIKASPHADPPSAFSAFPPCHPLPLGLHDTARVLP 1098  
DB 319 -----GD--TGAKGPPGVQ-----GPPGAGEGKRGARCEPCTG-----LPG 358  
QY 1099 PPTISNPPPLISSAKHPSVLIERQIGAISOGMSVOLHVPISEHAKAPVGPV-TNGLPLPMD 1157  
DB 359 PPG-----ERG-GPGSRGF-----PGADGVAGPKGAGERGSPGAG 394  
QY 1158 PKKLAPFSGVKQQLSPRGAQPPESLGVPTAQEASVLRGSTALGSPVGSITKIGISTRV 1217  
DB 395 PKG-----SP-GEAGRPEAGLPAGKLTGSPGS-----PGPDGKTGPPG---- 433  
QY 1218 PSDSAITYRGSITHGTTPADVLYKGTITRIIGEDSPSRDLRGREDSL-----PKGHVITYEGK 1273  
DB 434 PAGQ-----DGRP-----GPPGPPGA-RGOAGVMGFPKGAAGEPCK 470  
QY 1274 KGHVLSYEGM-----SVTQCKSKEDGRSSGPPHETAAPKRTYDMMGRVGRRAISSASIE 1328  
DB 471 AG-----ERGVPPGPGAVGPGAGKDGGAQGGP-GPAGPAGE-----RGQGGA-GSPGFQ 519  
QY 1329 GLMGRAPPP-ERHSPHLLKEQHHRIGSITQIGIPRSVIEAQEDYLRREKLLKREGTPPPP 1387  
DB 520 GLPGPAGPGEAGKPE-----QGVPGDLGAPGSPGARGERFPGERGVOGPP 567  
QY 1388 PPSRDLTEAYKQALGPLK-PAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPRP 1445  
DB 568 GPA-----GPRGANGPNDGAKG---DAG-----APCAP-- 594  
QY 1446 LKEGSITQCTP-----LKDYTGASTTGSKKHD-----VRSLSGSPGR-----TFP--PVH 1488  
DB 595 ---GS--QGAPGLQMPGERGAAGLPKPGKDRGDAGPKGADGSPGDRGLTGTPGPPG 649  
QY 1489 PLDMADARALERACEESLKRPGTASSGSGSIARGAPVIVPELCKPQSPITYEDHGA 1548  
DB 650 PAGAPD-----KGESGSPGAPGTGARGAPGDRGEPGP--GPAGFA--GP 692  
QY 1549 PFAGHLPRGSPVTMREPTPRLQEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPPI 1608  
DB 693 PGAD---GQCAKGEPPDAGAKD-----ACPPGAPGAPGPGPI 729  
QY 1609 -----SPYEHLRLRGVSDLYRSHIPLAFDPTSIPIRGIPLDAAAAYYLRHPLAPNPTYPHL 1664  
DB 730 GNVGAPGAKGARGAGP-----PGATGFPGAAGRVP-----PGPS-GNA 768  
QY 1665 YPPYLIRGYPDTAALLENROTIINDYITSOQMHNTATAMAQRADML-----RGL 1713  
DB 769 GPP-----GPPGAPGEGGKGRGE-----TGPAGRPEVGPVPPGPPAGEKGS 812  
QY 1714 SPRESSLALNAAAGPRGIIDLQVPHLPVLVPPT--PGPATAMDRALVLTAPQPFSSR 1771  
DB 813 PGADGPAGAPGTGPGGIAGQGVGLPCQGRGERFPGPLPGPSG-----PKQGSFGAS 867  
QY 1772 HSSSPLSPGPGTHLTKPTTTSSESRERDRDRDREREKSILTSTTTVEHAPIWRPOT 1831  
DB 868 GERGPPGMPGPLAGPGESGRE-----CAPGA 896  
QY 1832 EGSSSSSSSSGGGG-----GSSSRPASHAHQHSPISPRTQDALQORPSVLHNTCMKGI 1886  
DB 897 EGSPGRDGSFGAKGRGETGAPGFCAGXAGAPGVGP-----AGKSGD 941  
QY 1887 ITAVPSKPTVLRSTSTSSVPRPAATFPATHCPLGGTLGVYPTLMEPVLLPKAPRVA 1946  
DB 942 RGETGAPG-----AGPVGAPGARGP-----A 963  
QY 1947 RPERPRADTGH-----AFLAKPPAKSGLEPASSPSKSGSRPLRPPVPSGH 1991



Db 964 GPOGRGDKGETGEQGRGIKIHGRGFGSLQGGPPGPGSGQPGSGAGSP----- 1013  
Qy 1992 ATIAITPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQEELSLGVLHGSSYSP 2051  
Db 1014 -----AGPRGPGSAGAPGKDGUNGLPPI-----GPPGRGRGTG 1048  
Qy 2052 GVEPVSPVSSPLTHDKGLPKHLELDKSHLEGLRKPQPGVKLGGEAAHLPLRLPLPE 2111  
Db 1049 DAGVPGPGPG-----PPGPPGPPSAGDFSLP----- 1078  
Qy 2112 SOPSSPLQTAGVKGHQ-----RVVTLAHSIEVITQDYTR 2149  
Db 1079 -QP-----POBAKHGGRYRADDANVVRDRDLEVDTLKLSQOQIENIRSPGKR 1128  
Qy 2150 HIPOOLSAPLAPLYSFFCASCPLDLRRPPSDLYLPPDHCAPARGSPHSGGKRSP 2209  
Db 1129 KNPAR-----TC--RDLKCHSDW-----KSGEYWDIP 1154  
Qy 2210 NKTSLVGGGEGIEVPSPPEGMTEPGHRSAYVPLLYRDGEQTEPSRMG-----SKSPCN 2264  
Db 1155 NQ---GCNLDIAKVCNME-----TGTCVYP-----TOPSVAQKNWYISKPKD 1196  
Qy 2265 TSQPAFASKLTES-----NSAMVSKKQBINKLNTNHNREPEYNIS 2307  
Db 1197 KDRHVMFGESMTDGFQFEYGGGSDPADVAIQLTFLRLMSTEASQNTITYHCKNSVAYMDQ 1256  
Qy 2308 QPG 2310  
Db 1257 QTG 1259

## RESULT 9

US-07-741-940-7  
Sequence 7, Application US/07741940  
Patent No. 5352775

## GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERTS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07741,940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-07-741-940-7

Query Match 2.7%; Score 350.5; DB 1: Length 2842;  
Best Local Similarity 17.8%; Pred No. 1.1e-11;  
Matches 473; Conservative 357; Mismatches 1007; Indels 821; Gaps 109;

Qy 50 ASHLSPGSIQIP--QRRRPSLLSEFQ-----PGNERSQELHLRPSHSY 91  
Db 739 ANINSPGSSLPVHRKQKALEAELDAQHLSFTFNIDNLSPKASHRQKHQSLYGDY 798  
Qy 92 LPELGKSEMEIESKRPRLELLPDLLRPSLLATGQAGSEDLT---KDRSL----- 141  
Db 799 VFDTRNRDDNRSDFNTGNTMTVLSPLYLNTVLPSSSSRGSLDSSRSEKDSLERERGIG 858  
Qy 142 TGKLEPVPPSPPHDPELELVPPRLSKLEELIQNMDRVREITMVEQQISKKKKQOOLE 201  
Db 859 LGNYHPATEN-----PGTSSKRG-----QISTTAAQIAKV-----ME 891  
Qy 202 EEAAPPEPEKVPSPPIESKHSRLVQIYDNR--KKAEEAHRILEGLGPOVELPLYNQ 259  
Db 892 EVSAIHTSQEDRSSGSTTE-----LHCVTDERNALRSSAAH-----THSNTYN- 935  
Qy 260 PSDTROYHENIKINQAMRKLLIFYKRRNHARKQWKQFCORYDOLMEALEKKVERIENN 319  
Db 936 --FTKSENSNRTCSMPYAK---LEYKRSSN-----DSL-----NSVSSSDGY 972  
Qy 320 PRRRAKESKVRREYKEK-----QFP-EIRKQRELOERMQSRVQSGSGLSMSAARSE 369  
Db 973 GKRGOMKPSIESYSEDESCKFCYGYPADLAHKTHSANHMDNDGELDTPIYSLKYSYD 1032  
Qy 370 HEVSEIIDGLSEQENLEKQMLAVIPMLYDADQORIKFINMGLMADPMKVYKDRQVM 429  
Db 1033 -----EQLNSGRQ-----SPSONERWARPKHIIEDIEIKSQSQRSR 1068  
Qy 430 NMSWSE--KETFREKFMOHPKNFGLIASFLERKTVACVLYYYLTKKNNYKSLVRSY 487  
Db 1069 NOSTTYPVYTESTDDKHLKFPHPG-----QQECVSPYRSRGANGSETNVRGNSH 1118  
Qy 488 RRRGKSOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 541  
Db 1119 ---GINQNVQSQCQEDDYEDDKPTNYSERYSEEQHEEERPTNYSIKYNEEKRHVDQ 1175  
Qy 542 KEDLLKERTD-----DTSGEDN-DEKPAVASKGKRTANSQORRGRITRSHA 587  
Db 1176 IDYSLKYATIDIPSSOKQSFSSKSSGSSKTEHMSSENTSPSSNAKRONQLHPSSA 1235  
Qy 588 NEANSEEAITPOQSAELASMEINNESSRWTEEMETAKGELLEHGRNWSAIARMVSKTVS 647  
Db 1236 QSRSGQ----PQKAATCKVSSINQETIQTYCVEDTP-----ICFSRCSLSLS 1280  
Qy 648 QCKNFYFNKKRONLD--EILQOHLKMEK--ERNARKKKKAPAAASEAAFPVWVEDE 703  
Db 1281 SAEDEIGCNQITQEADSANTLQIAEIK-EKIGTRSAEDPVSEVPVAVSOH-----PRTSS 1334  
Qy 704 EMEASGVSGNEE-----EMVEEAALHASGNEVPGE-----CSGPATVNNSSDPTSPS 753  
Db 1335 RLQSSLSSESARHKAVEFSSGAKSPSKGAQTPKSPPEHYVQETPLMFSTRCTSVSSLDS 1394  
Qy 754 PHTAAKDTGQNGPKPPATLGADGP-----PGPPTPRRTSRAPIEPTPASEATGATP 808  
Db 1395 FESRSIASSVOSEPCSGMWGIIISPSDLDPDPSGOTMPPSRSK-----TP 1438



QY 809 PPAPSPSPVPPVKEEKEETAAAPPVEEGEQKPPAAEELAVDTGKA---EPPVKSE 865  
Db 1439 PP-PQTAQTKREVPKNK-----APTAEKRESGPKQAANVAQVQVLPDADLLHF 1490  
QY 866 CTEEAEG-----PAKGDAAEAATAGALKAEKKEGGSRATTA---KSSGAPQDS 915  
Db 1491 ATESTPDGFCSSSLSALSDEFFQOKVELRIMPVPVQNDNGNETESQPKESNQBK 1550  
QY 916 DSSATCSA---DEVDAEGD---KNRLSPRPSLLTPTGDPANASPOKPLDLKOLKOR 969  
Db 1551 EAEKTIIDSEKLLDDSDDDIEILEECIISAMPT-----KSSRAKAP-----AOT 1596  
QY 970 AAAIPPIQVTKVHEPPREDAAPTKPAPPAPPONILQPEDAPQPGSS-PRGKSRSPAP 1028  
Db 1597 ASKLPPPVARKPSQLPVYKLLPSQ-----NRLQPKHVSFTPGDDMPR-----1639  
QY 1029 PADKEAFAEAQKLCGDPCCWTSGLFPFVPPPREVIKASHAPDPSAFSTAPGCHPLPLGL 1088  
Db 1640 -----VYCVGEG-----TPINFSTA-----TSL 1656  
QY 1089 HDTARVLPRPPTISNPPLISSAKHPSVLERQIGAISQMSVOLHVPHYSEHAKAPVGPV 1148  
Db 1657 SDL-----TIESPP-----NELAAGEVGRGAQGEFEKRDITPT 1691  
QY 1149 TMGLPLMPDKKLAPFSGVYKQQLSPRGQAGPPESLGVP-----TAQEAASVLRGTALGSV 1203  
Db 1692 -----EGRSTDEAOGKTSSTVTPELDDNKAEEGDILAECINSAM 1731  
QY 1204 PG-----SITKGIPTSRVPSDAITVR-----GSITHGTPADVLYKGTITRI 1246  
Db 1732 PRGKSHKPPRVKIMDVQOASASSAPKNQLDGKKKPTSPVKPIQONTYRTVRK- 1790  
QY 1247 IGEDSPRLDRGREDSPKGVHYIEGKGVLSYEGGMSVTOCSKD-----GRSSSGPPE 1303  
Db 1791 -NADSKNNLNAERVS DNK-----DSKONLKNNSKDFNDKLPNNEDRVGSGFADSPHH 1844  
QY 1304 TAAPKRTYDMGRVGRRAISSAIEGL-----MGRAIPPERHSPHLKEQHHRGSI 1355  
Db 1845 -----YPIEGTPVCFNRNDSLSLDFDDDDVLSR-----EKAELRAKENKESEAKV 1893  
QY 1356 TOGIPRSVYEAQEDYLREAKLLKREGPPP-----PPPSROL-----TEAKTO 1400  
Db 1894 TSHTELTSNQQSANKTQAIKAPINRGQPKPILOKQSTFPQSSKDIPDGAATDEKLQNF 1953  
QY 1401 ALGPKLKPAHGLVATVCAKGRSTHEIPRELRT-----PELPLAPRLKEGSIT 1452  
Db 1954 AIENTPVCFSHNSLSLSLSDIOENNNKENEDIKETEPDQSGEFSKPOASGAPKSFV 2013  
QY 1453 QGTPLKYDTGASTG-----SKHDRVSLIGSPGR 1482  
Db 2014 EDTPVCFNRNDSLSLSIDSEDDLQECISSAMPKKKPSRLKGDNEKHSRNMGGILGE 2073  
QY 1483 TFPVPHPLDVMDARALERACVYESLSKSPGTASSGGSGIARGAPVIVPELCKPROSPIT 1542  
Db 2074 -----DLTLDLKDIOQPDSEHGLS---PDSENFWDKAIQEGANSIVSSL-----2114  
QY 1543 YEDHGAPFAGHLPR-----GSPVTREPTPRLOEGSLSSKASQDRK 1584  
Db 2115 ---HQAARAAACLSRQSSDSLSLSKSGISLGSFHL---TPDQERFTSNKG-----2163  
QY 1585 LUTSTPREIAKSPHSTVPEHHPIPSYEHLLRGV-SGVDLYRSHIPLAFDPTSIPIRGIP 1643  
Db 2164 ----PRILKPGKSTL-----ETKKIESESKIGKKGVKVS-----2196  
QY 1644 DAAAAYLPLHLPNTVPHLYPPVLYRCYPTATALENQTIINDYITSQOMHHNTATAM 1703  
Db 2197 -----LITG-----KVRNSNEISQOM-----2212  
QY 1704 AQRADMLRGLSPRESSLALNYAAGRPIDLSQVPHLPVLPVPTTGTATAMDRLAYLPT 1763  
Db 2213 -----KQPLQANMPSISRG-----RTMTHIFGV-----2235  
QY 1764 APQPFSSRHSPLSPGPGTHLTUKPTTSSSRERDRDRDRDREREKSILTTT-----1820

Db 2236 ----RNSSTSTSPVSKKGP-LKTPASKSPSEQ-----TATTSRPG 2272  
QY 1821 -----VEHAPITWPCTEOSSSGSSGGSSGGSSPASHSHAHQHSPISPRTDALQOR 1874  
Db 2273 AKPSVKSLSPLVARO-TSQIGSSKAPSRGSRDSTPSRPAQOPLSRPIQSPGRNSI---2328  
QY 1875 PSVLHNTGKGIITAVEPSKPTVLRSTSTSSPVVRPAATPPATHCPLGGVLDGVVP---1930  
Db 2329 -----SPGRNGI---SPPNKLSQLPRTSPSTA-----STKSSGSKMSTSPGRQM 2372  
QY 1931 ---TLMPEVLLKPEAPRVARPERPRADTGHAFIAKAPARSGLEP-----ASSPSKSGE 1980  
Db 2373 SQONLTQKGLSKNASSI-----PRSESASKGLNOMNGNANGANKVELSRMSTFKSSGE 2427  
QY 1981 ---PPLVPPVSGHATIAITPA-KNLAPHHASPDPPAPASADPHREKTSQKPSIQE 2035  
Db 2428 SDRSERPVLRQSTFTKEAPSTLRKLEESASFESLSSSRPASPTRSQAOT-PVLSPS 2486  
QY 2036 LELRSLGYHGSSVSPGVPVSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQCPVK 2095  
Db 2487 LPMDSLTH-SSVQACGWKLPNLSPTTEYNDGRPAKRDHDIARSHSESPSL---PIN 2541  
QY 2096 LEG--EAAHLPHLRPLPESQSSPLLQAPGVKGHQRVVTLAQHISEVITQDYTRHPQ 2153  
Db 2542 RSGTWKREHSKH-----SSSLPRVSTWRTGSSSSILSASSESEKAKSEDEKH---2590  
QY 2154 QLSAPLAPLYSPGASCPVLDLRRPPSDLYLPPPHGAPARGSPHSEGGKRPENKTS 2213  
Db 2591 -----VNSISGTK-----QSKENQVSAKGTWRKIKENEFSPNTNSTS 2626  
QY 2214 --VLGGEGDIE-----PVSPEGMTEPGHSRSAYVPLLYRDGEOTEPSRMGSKSPGNT 2265  
Db 2627 QIVSSGATNGASKTLIYQMAVAKSTEDVWVRIEDCI-----NNPRSRSTGNT 2678  
QY 2266 SOPPAFFSKLTESNAWVSKKQKINKLTHNRNEPEYNISQPGTEIFNMPAITGTGLM 2325  
Db 2679 --PP-----VIDSVSEKANPNIKSKDNQAKNV-----GNGSV 2710  
QY 2326 TYRSQAVQEHASTNMGLEALIRKALMGKVDWEESPLSANAFNPLNASASLPAAPIA 2385  
Db 2711 PMRTVLENLNLSFIQVDAPDQKGTIKPCQNNPVVSETN-----ESSIVERTPFSS 2763  
QY 2386 ADGRSDHTLTPGSGGKAKVSG---RPSRKAKSPAPGLASGDRPPSVSVHSEGDENR 2442  
Db 2764 SSS-SKH---SSPSGTVAARVTFNPNPSRKSAD-----STSRPSQIP-----2805  
QY 2443 TPLTRVWEDRPSSAGST 2460  
Db 2806 TPVNNT-KKRSKTDST 2822

## RESULT 10

US-08-289-548A-7

; Sequence 7, Application US/08289548A

; Patent No. 5648212

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

; APPLICANT: CARLSON, MARY

; APPLICANT: GRODEN, JOANNA

; APPLICANT: HEDGE, PHILIP J.

; APPLICANT: JOSLYN, GEOFF

; APPLICANT: KINZLER, KENNETH

; APPLICANT: MARKHAM, ALEXANDER F.

; APPLICANT: NAKAMURA, YUSUKE

; APPLICANT: THLIVERIS, ANDREW

; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

; NUMBER OF SEQUENCES: 102

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner &amp; Allegretti, LTD

; STREET: 1001 G Street, NW







Qy	1483	TFPPVPHLVMDADARALERACRYEESLSKBPCTASSSGGSIARGAPVIVYBELGRPQSGPLT	1542
Db	2074	-----DUTLDKDIORPDSEHGLS--PDSENFDMAKIOEGANSIVSSL-----	2114
Qy	1543	YEOHGAFFAGHLPR-----GSPVTWREPTPLQEGSISSSSKASODRK	1584
Db	2115	--HQAAMACLRSQASSDSLSLKSGISGSPFHL--TPQOEKPTSNKG-----	2163
Qy	1585	LTSTPREIAKSHUSTVPEHHHPISPYEHLRGV--SGVDLYRSHIPLAFDPTSIPIRGIP	1643
Db	2164	-----PRILKPGEKSTL-----ETKKIESEKINGKGGKVKYS-----	2196
Qy	1644	DAAAAYVPHLAPNPTYPHLYPPVPIRGYPDTALENROFIINDYITLSQMHHTATAM	1703
Db	2197	-----LITG-----KVRNSEISGOM-----	2212
Qy	1704	AQRMDLRGLSPRESSALNYAAGPRGIDLSQVPHPLVLPVPTPGTATAMDRLAYLPT	1763
Db	2213	-----KOPLOANMPSISRG-----RTMIHIPGV-----	2235
Qy	1764	APQFSSRHSSPLSPGGTHLTKTPTTYSSEREDRDRDRDREREKSLTSTTT--	1820
Db	2236	-----RNSSSTSPVKKGPP-LKTFASKSPSEGO-----TATTSPRG	2272
Qy	1821	-----VEHAPITWRPTEQSSGSSGGGSSRRPASHAHQHSPISPRQDALOQR	1874
Db	2273	AKPSVSELSPVARQ--TSQIGSSKAPSRSGRSDTSPSPAQQPLSRPIOSPGENSI--	2328
Qy	1875	PSVLNHTGKGIITAVEPSKPTVLBSTSTSPVRPAATFPPTHACPIGGTLDGVYP--	1930
Db	2329	-----SPGRNGI--SPPNKLSQLPRTSSPSTA-----STKSSGSKMYSITSPGRQM	2372
Qy	1931	-----TLMPEVLLKPEARVARPERPRADTGHAFKAPPARSGLEP-----ASSPKSGE	1980
Db	2373	SQONLTKQGLSKNASSI-----PRSEASGGLQNMNGNCAKVKELRMSSTKSSGE	2427
Qy	1981	-----PRPLVPVSGHATIARTA-KNLAPHASDPDPAPPASADPHREKTQSKPFSIQE	2035
Db	2428	SDRSERPVLVROSTFTKEAPSPPTLRKLEESASFESLSPSRPASPTRSQAQT--PVLSPS	2486
Qy	2036	LELRSLGYHGSSVPBGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQCPVK	2095
Db	2487	LPDMSLSTH--SSVQAGGWKRLPNLSPITIEDNCRPAKRHDIAHSHSESPSRL-----	2541
Qy	2096	LGG--EAAHLPHLRPLPESQSPSSPLQATAPGVKGHVVTFLAQIHSEVITQDYTRHHQP	2153
Db	2542	RSGTWKRHSKH-----SSSLPRVSTWRTGSSSSSILSASESKAKSEDEKH--	2590
Qy	2154	QLSAPLPAPLYSPFGACPVLDLRPPSDLYLPDPDHGAPARGSPHSEGGKRSPENKTS	2213
Db	2591	-----VNSISGTK-----QSKENOVSAKGTWRKIKENEFSPTNSTS	2626
Qy	2214	--VLGGEDGIE-----PVSPPEGMTBPGHSRAVYPLLVRDCEQTEPSRMGSKSPGNT	2265
Db	2627	QTVSSGATNGASKTLIYMAAFVSKTEVDVVRVEDCPI-----NNPFRSGRSPGTNT	2678
Qy	2266	SQPPAFPSKLTESNAWVSKKQEIKNKLNTHNRNEPYNISQGTETFNMPAITGTGLM	2325
Db	2679	--PP-----VIDSVSEKANPNIKDSKONQAQONY-----GNGSV	2710
Qy	2326	TYRSQAOVBEHASTNMCLEAIBKALMGKYDQWESPPLSANAFNPLNASLASLPAAMPITA	2385
Db	2711	PMRTVGLNRLNSFTQVDAPODKGTETPKQGNPNVPVSETN-----ESSIVERTPFSS	2763
Qy	2386	ADGRSDHRTLSFGGGKAKVSG--RPSRRKAKSPAGLASGDRPPSPVSVHSEGDGNRR	2442
Db	2764	SSS-SKH--SSPSGTVAARVTFPNINPSPRKSSAD----STPSARPSQIP-----	2805
Qy	2443	TPUTNRWEDRPSAGST	2460
Db	2806	TPVNNNT-KRKDSKTDST	2822

RESULT 11  
 US-08-452-654-7  
 : Sequence 7, Application US/08452654  
 : Patent No. 5691454  
 : GENERAL INFORMATION:  
 : APPLICANT: ALBERTSEN, HANS  
 : APPLICANT: ANAND, RAKESH  
 : APPLICANT: CARLSON, MARY  
 : APPLICANT: GRODEN, JOANNA  
 : APPLICANT: HEDGE, PHILIP J.  
 : APPLICANT: JOSLYN, GEOFF  
 : APPLICANT: KINZLER, KENNETH  
 : APPLICANT: MARKHAM, ALEXANDER F.  
 : APPLICANT: NAKAMURA, YUSUKE  
 : APPLICANT: THLIVERIS, ANDREW  
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 : NUMBER OF SEQUENCES: 94  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Banner, Birch, McKie & Beckett  
 : STREET: 1001 G Street, NW  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: USA  
 : ZIP: 20001-4598  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/452,654  
 : FILING DATE: 25-MAY-1995  
 : CLASSIFICATION: 536  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/741,940  
 : FILING DATE: 08-AUG-1991  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Kagan, Sarah A.  
 : REGISTRATION NUMBER: 32,141  
 : REFERENCE/DOCKET NUMBER: 1107.035574  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 202-508-9100  
 : TELEFAX: 202-508-9239  
 : INFORMATION FOR SEQ ID NO: 7:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2842 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : ORIGINAL SOURCE:  
 : ORGANISM: Homo sapiens  
 : IMMEDIATE SOURCE:  
 : CLONE: APC  
 : US-08-452-654-7

Query Match 2.7%; Score 350.5; DB 1; Length 2842;  
Best Local Similarity 17.8%; Pred. No. 1.1e-11;

[illegible]











Db 1492 ATSTPDGFCSSLSALSLEDEFFIQKQVELRIMPPVOENDNGNETESBQKESNENQEK 1551  
Qy 916 DSSATCSA--DEVDAEGCD---KNRLLSPRPSLLTPTGDPANASQKPLDLKQLQOR 969  
Db 1552 EAETIDSEKDLDDDDDDIILEECIISAMPT-----KSRKAKKP-----AQT 1597  
Qy 970 AAAIPIQVTKVHEPPREDAAPTAPAPPAPPPQNLQOPESDAPQPGSS--PRGKRSRSPAP 1028  
Db 1598 ASKLPVPVARKSOLPVVLLPSQ-----NRLOQKHVSFTPGDDMPR-----1640  
Qy 1029 PADKFAFAEAQKLPDPCWTSGLFPVPPREVIKASPHADPDSAFSYPAGHPLPLGL 1088  
Db 1641 ----VYVEG-----TPINFSTA-----TSL 1657  
Qy 1089 HDTARVLPPTTISNPPPLISSAKHPSVLERQIGALSQMSVOLHVYSEHAKAPVGPV 1148  
Db 1658 SGL-----TIESPP-----NELAAGEGVGGAOSGEFEKRDITPT 1692  
Qy 1149 TMGLPLMPDKKLAPFSGVKQBLSPRGQAGPPESLGV-----TAQASVLRGTALGSV 1203  
Db 1693 -----EGRSTDEAQQGKTSVVTPELDDNKAEGDILAEICINAM 1732  
Qy 1204 PGG-----SITKGPSTRVPSDSAITVR-----GSITHGTADVLYKGTITRI 1246  
Db 1733 PKGSHKPFVKKIMDQVOOASASAPKNQDKKKKPTSPVKPIPONTEYRTVRK- 1791  
Qy 1247 IGEDSPRLDRGREDLPGHVIYEGKKHVLSEYGGMSVTCCKED---GRSSSGPHE 1303  
Db 1792 -NADSKNNLNAERFSDNK-----DSKKONLKNNSKDNKLPNNEDRVGFSFADSPHH 1845  
Qy 1304 TAAKRTYDMMGRVGRATISSAIEGL-----MGRAPPERHSPHLLKEQHIRGSI 1355  
Db 1846 -----YTPISGTYCFSRNDSLSLDFDDDDVLSR-----EKALRKAKENKESEAKV 1894  
Qy 1356 TQGPISVVEAQEDYLREAKLLKREGTTPP-----PPSRDL-----TEAYTKQ 1400  
Db 1895 TSHTELTSNQQSANTQATAIKQPINRGOPKPILOKQSTFPQSSKIDIPRGAATDEKLQNF 1954  
Qy 1401 ALGPLKPAHGLVATVKEAGRSIHEPRELRT-----PELPLAPRLKEGSIT 1452  
Db 1955 AIENTPVCFSHNSLSLSDIDDOENNNENEFIKETEPDQSGEFSKFOASAYAPKSFHV 2014  
Qy 1453 QGTPKYDTGASTG-----SKKHVRSLLIGSPGR 1482  
Db 2015 EDTPVCFSRNSSLSDISEDDLQECISSAMPKKKPKSRLLKGDNEKHSRPMGILGE 2074  
Qy 1483 TPPPVPDVMADARALERACVYESLSKSPGTASSGSGSIARGAPVIVPELCKPQSPILT 1542  
Db 2075 -----DLTLDLKDIOQPDSEHGLS--PDSENFWDKAIQEGANSIVSSL-----2115  
Qy 1543 YEDHGAPFAGHLPR-----GSPVTWREPTPRLOEGSSSSKASQDRK 1584  
Db 2116 ---HQAAAACLSRQASDSDSILSKSISLGSFHL---TPDQEEKPFTSNKG-----2164  
Qy 1585 LTSTPREIAKSPHSTVPEHHHPHISPYEHLRGV--SGVDLYRSHIPLAFDPTSPRGIPL 1643  
Db 2165 ----PRILKPGKSL-----ETKKIESSEKRGKGVKVS-----2197  
Qy 1644 DAAAAYLPRHLAPNTPVPLYPVPLIRGYPTDALENQRTIINDYITSQMHHTATAM 1703  
Db 2198 -----LITG-----KVRNSEISGOM-----2213  
Qy 1704 AQRADMLRGLSPRESSLALNAAAGPRGIIDLQVPHLPVLPPTPTGTATAMDRLAYLPT 1763  
Db 2214 -----KQPLQANPISRG-----RTMHIHGV-----2236  
Qy 1764 APQPFSSRHSSPLSPGGPHTLTKPTTTSSSERDRDRDRDREREKSLTSTTT---1820  
Db 2237 ----RNSSSSTSPVSKGKPP-LKTPASKSPSEGQ-----TATTSRPG 2273  
Qy 1821 -----VEHAPLWRPCTEQSSGSSSGSGSGSSRSPASHAHQHSPISTPTQDALQOR 1874  
Db 2274 AKPSVKSELSPVARO-TSQIGGSKAPSRSGSRDSTPRPAQQLSRPIQSPGRNSI---2329

Qy 1875 PSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVRAAATFPATHCPLGCTLDGVYP----1930  
Db 2330 -----SPGRNGI---SPPNKLSQLPRTSSPSTA-----STKSSGSKMSYTPSGROM 2373  
Qy 1931 ---TLMPEVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEP-----ASSPSKSGSE 1980  
Db 2374 SOONLTKTGSLKSNASSI-----PRSESASKGLNOMNGNGANKKVELSRMSTKSSGSE 2428  
Qy 1981 ----PRPLVPPVSGHATIAITPA-KNLAPHASPPPPAPPASADPHREKTQSKFSTQIE 2035  
Db 2429 SDRSERPVLRQSTRIKEAPSTPLRKLEESASFESLSPSSRPASPTRSOAQT-PVLSPS 2487  
Qy 2036 LELRSLGVHSGSYSEGEVPVSPSPSLTHDKGLPKHLEELDCKHLECELPRKQPGPVK 2095  
Db 2488 LPDMSLSTH-SVQAGGWRKLPPLNLSPTIENYNDGRPAKRDHARSHESPSL---PIN 2542  
Qy 2096 LGG--EAAHLHLRLPLPESQSSSPLLQATAPGVKGHQRVVTLAQHISEVITQDYTRHHQ 2153  
Db 2543 RSGTWKREHSHK-----SSSLPRVSTWRTGSSSILSASSESEKAKSEDEKH---2591  
Qy 2154 QLSAPLAPLYFPGASCPLDLRRPPSDLYLPDPHGAPARGSPHSEGGKRSPPENKTS 2213  
Db 2592 ----VNSISGTK-----QSKENOVSAKGTWRKIKENEFSTNSTS 2627  
Qy 2214 --VLGGGEDGIE-----PVSPPEGTEPHGHSRAVYPLYRDGEQTEPSRMGSKSPGNT 2265  
Db 2628 QTVSSGATNGAESKTYLIYOMAPAVSKTEDVWVRIEDCPT-----NNPSRGRSPTGNT 2679  
Qy 2266 SQPPAFFSKLTESNAMYKSKQKQKINLTHNRNEPYNISQPGTEIFNMPAITCTGLM 2325  
Db 2680 --PP-----VIDSVSEKANPNIKOSKDNQAKQNV-----GNGSV 2711  
Qy 2326 TYRQAOVQEHASTNMGLEAIRKALMGKYDQWEEPPPLSANAFNPLNASASLPAAMPITA 2385  
Db 2712 PMRTVGLNRLNLSFIQVDAPQKTEIKPGQNNPVVSETN-----ESSIVERTPFSS 2764  
Qy 2386 ADGRSDHPLTSPGGGKAKVSG---RPSRRAKSPAPGLASGRDPPSVSVHSEDCNRR 2442  
Db 2765 SSS-SKH--SPSGTVAARVTPFNTPNPSRKSAD----STSRPSQIP-----2806  
Qy 2443 TPLTNRVNREDRPSAGST 2460  
Db 2807 TPVNNNT-KKRDSTKTDST 2823

## RESULT 13

US-08-452-655B-7  
; Sequence 7, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk



















Qy	866	CTEAEEG-----PAKCKDAEAEATAEGALKAEKEGGSGRATTA---KSSGAPQDS	915
Dy	1492	ATESPTDGFSCSSLSALSDEPFFIKQDVLRIMPVQENDNGNETESEQPKESENQOEK	1551
Qy	916	DSSATCSA--DEVDEAEGD---KNRLLSRPSSLITFTGDPRNASSPOKPLDLKLQKQR	969
Dy	1552	EAKTIDSEKDLLDDSDDDIIEILECISAMPT-----KSRRAKKP-----AQT	1597
Qy	970	AAAIPIQTIVKHEPPREDAAPTKPAPPQQPNLOPESDAPQOQGSS--PRGKSRSFAP	1028
Dy	1598	ASKLPPPVARPKPSOLPVYKLLFSQ-----NRLOPQKHVSFTPGDDMPR-----	1640
Qy	1029	PADKEAFAAEAQLPGDPCCWTISGLFPVPREVTKASPHADPSAISAYPPGHLPPLGL	1088
Dy	1641	----VYCVEG-----TTTTTTT-----TTFNSTA-----TSL	1657
Qy	1089	HDTARVLP RPPTISNNPLLISAKHPVLERQIGAISOGMSVOLHVHYSEHAKAPGPV	1148
Dy	1658	SDL-----TIESPP-----NELAAGEVGVRGGQAQSGEFKROTIPRT	1692
Qy	1149	TMGFLPMDPKLIAPESGVKYQQLSPRGQAGPPELSGVP-----TAQEASVLRGTALGSV	1203
Dy	1693	-----EGRSTDEAQQGKTSSVTIPLELDNKAEEDILAEBCINSAM	1732
Qy	1204	PGG-----SITKGTPSTRVPSDSNIYR-----CSIHTGTPADVILKGTITRI	1246
Dy	1733	PKGSHKPFVRVKIMDOVOQASASSAPKNQLOGKKKKPTSPVKPIPONTEYTRVRYRK-	1791
Qy	1247	IGEDSPRLDRGEDSLPKGHVIYEKKGHVLSYBGGMSVTOCSKED---GRSSSGPPHE	1303
Dy	1792	-NADSKNNLNAEVFSDNK-----DSKQNKLNNKSFDFNDKLPNNEDRVRGSAFDSPHH	1845
Qy	1304	TAAPKTYTDMGEVRGATISSASIELG-----MGRAIPERRSHPHLUKEQHIIIRGSI	1355
Dy	1846	-----YTPIEGTPTCYCFSRNDSLSSLFDDDDVDLSR-----EKAEULKAKENKESAKV	1894
Qy	1356	TQGISPYVEAQDYLRREAKLLKREGTPPP-----PPPSROL-----TEAVKQTQ	1400
Dy	1895	TSWTELTSNQOSANKTQAJAKOPINRGOKPILQKQSTFPSSDKDIIDPGAATDEKIQNF	1954
Qy	1401	ALGPLKLKPAHEGLVATVKAERSIHETPRELRHT-----PELPLAPRPLKEGSIT	1452
Dy	1955	AIENTPVCFSHNSSLSLSDIOENNKENEPKETETPDPSQGEPSKPQASGYAPKSFHV	2014
Qy	1453	QCTPLXYDCASTTG-----SKHDVRSILGSPGR	1482
Dy	2015	EDTPCVCFSRNSSLSLUSIDESDDLQEOCISSAMPKKKKPKSLKGNKDNKHPNNMGTLGE	2074
Qy	1483	TFPPVPHPLDVADARALERACYEESLKRGPTASSSGSIAARGAPVTVPELGKPRQSPLT	1542
Dy	2075	-----DLTLDLKDIOQPDSEHGLS--POSENFDWKAIQEGANSIVSSL-----	2115
Qy	1543	YEDHGAPFAGHLPR-----GSPVTVMREPTPRLOEGSLSSSKASQDRK	1584
Dy	2116	---HQAAAACLSRQASSDSLSILSKSIGSLGSPFHL---TPDQEEKPFTSNKG-----	2164
Qy	1585	LTSPTPREIAKSPHSVTPEHHHPISPYEHLRGV--SGVDLYRSHIPLAFOPTSIPRCIPL	1643
Dy	2165	---PRILAPGEKSTL-----ETKKIESEKGKGGKKVYS-----	2197
Qy	1644	AAAAAYVLPRLHAPNTYPHLYPPYLIRGYPDTALENROTIINDYITSQMHHNTATAM	1703
Dy	2198	-----LTTC-----KVRSNSELISGOM-----	2213
Qy	1704	AQRADMLRGLSPRESSLALNYAAGRPGIIDLSQVPHLPVLVPPPTGTPATAMDRLAYLPT	1763
Dy	2214	-----KQPLQANMPSISRK-----RTMIHIFGV-----	2236
Qy	1764	APQFSSRHSSSPLSPGGTHUTKTFTTTSSERERDRDRDREREKESILSTSTTT---	1820
Dy	2237	---RNSSTSSTSPVSKKGGP-LKTPASKSPSEGO-----TATTSPRG	2273

[illegible]

Search completed: September 8, 2001, 14:35:09  
Job time: 17211 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 09:57:13 ; Search time 70.16 seconds  
(without alignments)  
2732.773 Million cell updates/sec

Title: US-09-522-753-5  
Perfect score: 13215  
Sequence: 1 MSGSTQLVAQTWRATEPRYP.....WDEPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7871	59.6	1495	2 S60255	transcription co-r
2	4187	31.7	2453	2 S60254	nuclear receptor c
3	1272	9.6	1047	2 T46489	hypothetical prote
4	574	4.3	5327	2 T13564	microtubule-associ
5	552	4.2	2649	2 T51023	hypothetical prote
6	526	4.0	2142	2 B35098	MHC class III hist
7	520	3.9	2187	2 T30826	nascent polypeptid
8	512	3.9	5762	2 A41819	proline-rich pepti
9	508.5	3.8	1872	2 S36152	MHC class III hist
10	504	3.8	3942	2 T42730	Bassoon protein -
11	503.5	3.8	7962	2 T38346	elastic titin - hu
12	503	3.8	1870	2 S37671	MHC class III hist
13	502.5	3.8	2715	2 T13049	eyelid - fruit fly
14	498	3.8	3938	2 T42761	Bassoon protein -
15	492.5	3.7	2774	2 A43359	microtubule-associ
16	477	3.6	2282	2 T42717	DNA-binding protei
17	470.5	3.6	4957	2 T03455	ALR protein - huma
18	470.5	3.6	5262	2 T03454	ALR protein - huma
19	445.5	3.4	1952	2 T48814	hypothetical prote
20	436	3.3	1151	2 T18535	high molecular mas
21	424.5	3.2	2464	1 QRMSPI	microtubule-associ
22	419	3.2	5170	2 T15348	hypothetical prote
23	416.5	3.2	1791	2 T02345	hypothetical prote
24	412.5	3.1	1226	2 S15053	hypothetical prote
25	407.5	3.1	1464	1 CGHUI5	collagen alpha 1(I
26	406.5	3.1	2364	2 A56577	microtubule-associ
27	404	3.1	1274	2 T16251	hypothetical prote
28	403	3.0	1110	2 T15116	NF-180 - sea lamp
29	397	3.0	1670	1 CGHU3B	collagen alpha 3(I

30	395	3.0	1520	2 T00273	hypothetical prote
31	394	3.0	2783	1 A41948	alpha-fetoprotein
32	391	3.0	3924	2 S37431	ankyrin 2, neurona
33	389.5	2.9	1188	2 S49315	extensin-like prot
34	389.5	2.9	3869	2 A48205	All-1 protein +GRE
35	389	2.9	1320	2 JC5630	TCOF1 protein - mo
36	383.5	2.9	3968	2 A44265	trithorax homolog
37	383	2.9	2944	2 A54849	collagen alpha 1(V
38	377	2.9	1691	1 S22917	collagen alpha 5(I
39	375	2.8	1006	2 T42731	atrophin-1 related
40	374.5	2.8	2722	2 T20532	hypothetical prote
41	373.5	2.8	1184	2 S50832	atrophin-1 - human
42	373.5	2.8	1494	2 T14355	protein-tyrosine-p
43	373.5	2.8	2938	2 T30249	cell proliferation
44	373	2.8	1763	2 S16366	collagen alpha 2(I
45	372.5	2.8	1173	2 T31421	C-terminal domain-

ALIGNMENTS

RESULT 1  
S60255  
transcription co-repressor SMRT - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 05-Nov-1999  
C:Accession: S60255  
R:Chen, J.D.; Evans, R.M.  
Nature 377, 454-457, 1995  
A:Title: A transcriptional co-repressor that interacts with nuclear hormone receptors  
A:Reference number: S60255; MUID:96008552  
A:Accession: S60255  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1495 <CHE>  
A:Cross-references: EMBL:U37146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match		59.6%	Score 7871;	DB 2;	Length 1495;
Best Local Similarity		100.0%	Pred. No. 3.1e-269;		
Matches 1487;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1031	DKEAFAAEAAQKLPDPPCWTSGLPFPVPPREVIKASPHADPPSAFSAFVAPGHPPLGLHD	1090		
DB	9	DKEAFAAEAAQKLPDPPCWTSGLPFPVPPREVIKASPHADPPSAFSAFVAPGHPPLGLHD	68		
QY	1091	TARPVLPRPPTTISNPPLISSAKHPSVLERQICAIISQGSVQLHVPYSEHAKAPVGPVTM	1150		
DB	69	TARPVLPRPPTTISNPPLISSAKHPSVLERQICAIISQGSVQLHVPYSEHAKAPVGPVTM	128		
QY	1151	GLPLPMDPKKLAPFSGVKQOLSPRQAGPPESLGVPTAQEASVLRGTALGSPVGGISITK	1210		
DB	129	GLPLPMDPKKLAPFSGVKQOLSPRQAGPPESLGVPTAQEASVLRGTALGSPVGGISITK	188		
QY	1211	GIPSTRVPSDSAITYRGSIHTGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVY	1270		
DB	189	GIPSTRVPSDSAITYRGSIHTGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVY	248		
QY	1271	EKKKGHVLSYEGGMSVTCQSKEDGRSSSGPPHETAPKRTYDMMEGRVGRASISASIEGL	1330		
DB	249	EKKKGHVLSYEGGMSVTCQSKEDGRSSSGPPHETAPKRTYDMMEGRVGRASISASIEGL	308		
QY	1331	MGRAPIPERHSPHLKEQHHRIGSIQTGIPRSVVEAQDYLRREKLLKREGPPPPPPPS	1390		
DB	309	MGRAPIPERHSPHLKEQHHRIGSIQTGIPRSVVEAQDYLRREKLLKREGPPPPPPPS	368		
QY	1391	RLTEAYKTAQALGPLKPKPAHEGLVATVKEAGRSIHEIPREELRHTPELPAPRPLKEGS	1450		
DB	369	RLTEAYKTAQALGPLKPKPAHEGLVATVKEAGRSIHEIPREELRHTPELPAPRPLKEGS	428		
QY	1451	ITQGTPLKYDTGASTTGSKKHDVRSVLIGSPGRTFFPVHPPLDVNMADARALRACYEESLKS	1510		
DB	429	ITQGTPLKYDTGASTTGSKKHDVRSVLIGSPGRTFFPVHPPLDVNMADARALRACYEESLKS	488		



Qy	1511	RPGTASSGGSIARGAPVIVPELGKPRQSPPLYEDHGAPFAGHLLPRGSPVVTMRPTPLQ	1570
Db	489	RPGTASSGGSIARGAPVIVPELGKPRQSPPLYEDHGAPFAGHLLPRGSPVVTMRPTPLQ	548
Qy	1571	EGSLSSKASQDRKLNSTPREIAKSPHSTVPEHHPHPTSPYEHLRLRGVGVDLVRSHIPL	1630
Db	549	EGSLSSKASQDRKLNSTPREIAKSPHSTVPEHHPHPTSPYEHLRLRGVGVDLVRSHIPL	608
Qy	1631	AFDPTSPRGIPIDAAAAYLPHRLAPNTPYPLHYIRGYPDTAALENROQTIIINDYI	1690
Db	609	AFDPTSPRGIPIDAAAAYLPHRLAPNTPYPLHYIRGYPDTAALENROQTIIINDYI	668
Qy	1691	TSQOQHINTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPVPTPGT	1750
Db	669	TSQOQHINTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPVPTPGT	728
Qy	1751	PATAMDRLAVLPTAPQFSSRRSSSPLSPGGPTHLTKTPTTSSSERERDRDRDRDRER	1810
Db	729	PATAMDRLAVLPTAPQFSSRRSSSPLSPGGPTHLTKTPTTSSSERERDRDRDRDRER	788
Qy	1811	EKSLITSTTTVEHAPITWRPGTEQSSGSSGSSGGGGSSRRPASHSHAHQHSPISPRQTQDA	1870
Db	789	EKSLITSTTTVEHAPITWRPGTEQSSGSSGSSGGGGSSRRPASHSHAHQHSPISPRQTQDA	848
Qy	1871	LOQRPSVLHNTGKGIITAVEPSKPTVLIRSTSTSPVPAATFPATHCPLGGTLDGVYP	1930
Db	849	LOQRPSVLHNTGKGIITAVEPSKPTVLIRSTSTSPVPAATFPATHCPLGGTLDGVYP	908
Qy	1931	TLMEPVLLPKEAPRVARPERPRADTGHAPLAKPPARSGLGPASSPKSGSEPRPLVPVPSG	1990
Db	909	TLMEPVLLPKEAPRVARPERPRADTGHAPLAKPPARSGLGPASSPKSGSEPRPLVPVPSG	968
Qy	1991	HATITARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQELELSRIGYHGSSTSP	2050
Db	969	HATITARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFSIQELELSRIGYHGSSTSP	1028
Qy	2051	EGVEPVSSPSLTHDKGLPKHLELDKSHLEGELRKPQCPVKLKGEEAAHPLHLRPLP	2110
Db	1029	EGVEPVSSPSLTHDKGLPKHLELDKSHLEGELRKPQCPVKLKGEEAAHPLHLRPLP	1088
Qy	2111	ESQSSSPILLQTAGVKGHQRVVTLAQHILSEVITODYTRHHPOQLSAPLPAPLYSFFGAS	2170
Db	1089	ESQSSSPILLQTAGVKGHQRVVTLAQHILSEVITODYTRHHPOQLSAPLPAPLYSFFGAS	1148
Qy	2171	CPVLDLRRPDSLYLPPDPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGIEPVSPPEG	2230
Db	1149	CPVLDLRRPDSLYLPPDPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGIEPVSPPEG	1208
Qy	2231	MTEPGHSRSAYVPLYLRDGEQTEPSRMGSKSPGNTSQPPAFSKLTSNSAMVSKKQEI	2290
Db	1209	MTEPGHSRSAYVPLYLRDGEQTEPSRMGSKSPGNTSQPPAFSKLTSNSAMVSKKQEI	1268
Qy	2291	NKKLNTHNREPEYNTSQPCTEIFNNPATITGTLMTYRSQAVQEHASTNMGLEAIRKAL	2350
Db	1269	NKKLNTHNREPEYNTSQPCTEIFNNPATITGTLMTYRSQAVQEHASTNMGLEAIRKAL	1328
Qy	2351	MGYDQWEEPPILSANAFNPLNASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPS	2410
Db	1329	MGYDQWEEPPILSANAFNPLNASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPS	1388
Qy	2411	SRAKSPAGCLASDRPPSVSVHSEGCNRRPTLNTNWEDRSPSSAGSTPPFPYNPLIMR	2470
Db	1389	SRAKSPAGCLASDRPPSVSVHSEGCNRRPTLNTNWEDRSPSSAGSTPPFPYNPLIMR	1448
Qy	2471	LQAGVMASPPPGPLPAGSGPLAGPHAMDEEKPILCSQYETLSDSE	2517
Db	1449	LQAGVMASPPPGPLPAGSGPLAGPHAMDEEKPILCSQYETLSDSE	1495
RESULT	2		
S60254			
nuclear			
receptor			
co-repressor			
N-CoR			
- mouse			

C:Species: Mus musculus (house mouse)  
C>Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 08-Oct-1999  
C:Accession: S60254  
R:Hoerlein, A.J.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa Nature 377, 397-404, 1995  
A>Title: Ligand-independent repression by the thyroid hormone receptor mediates a negative transcriptional response to retinoic acid  
A:Reference number: S60254; MUID:96008539  
A:Accession: S60254  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-2453 <HOE>  
A:Cross-references: EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g1022718

	Query Match	31.7%	Score 4187;	DB 2;	Length 2453;
	Best Local Similarity	41.0%;	pred. No. 1e-139;		
	Matches 1103;	Conservative 348;	Mismatches 796;	Indels 444;	Gaps 1
Qy	16	EPRYPHSLSPVOIARHTDVGLELQ--HHSRDYASHLSPGSIIPQRPRPSLLSEFQ	73	:	:
Db	17	QSRYPSHVQVTFPSARHQGFAPVDYRSSHLEYSAQLQQOQQOOLRRPSSLSEFH	76	:	:
Qy	74	PGNRSQELHLRPESHVSLPELGKSEMFEISKRPRLLEPDPLLR-----PSSLATGQ	128	:	:
Db	77	PGSDRPQE--RRSGVEQHPGPSPVDHDSLESKRPLEQVSDSHFQRTISAVALPLVHT-L	133	:	:
Qy	129	PAG---SEDLPDKRSLTGKLE-PVSPSPPHTDDELEVLPRLSKEELIQNDVRDREIT	184	:	:
Db	134	PEGURSSANAKDPAFGVKYHEAPSPISGQPCGGDDONASPKLSKEELIQMSDRVDRIA	193	:	:
Qy	185	MVEQISKLKKKQOLEBEAAKPPEKVPSPPTIESKHSRLVLQIIYDENKKKAEEAHRI	244	:	:
Db	194	KVEQILKLLKKKQOLEBEAAKPPEKVPSPVPEQKHSRVIIYDENKKKAEEAHI	253	:	:
Qy	245	LEGIGPVVELPLYNQPSDTROYHENIKINQAMRKLLIIFYRRNHARKQWKQFCORYDQ	304	:	:
Db	254	FEGGLGVVELPLYNQPSDTKYHENIKINQVMRKLLIFFRRNHARKQBOKICORYDQ	313	:	:
Qy	305	LMEALEKVERIENPPRRRAKESKVREYEQFPEIRKORELQERMOSRGVORGSGLSMS	364	:	:
Db	314	LMEAEWVKVDRIENPPRKAKESKTREYEQFPEIRKORQOEERFQ-RVGQRGAGLSAT	372	:	:
Qy	365	AARSEHEVSEIIDLSEQLENLEQMROLAVIPPMLYDADQOIRKIFNMNGLMADPMKYK	424	:	:
Db	373	IARSEHETSETIIDLSEQENNEKMROLSVTPPMFMFAEQSRVFFINMGMLMEDPMKYK	432	:	:
Qy	425	DROYNNMSEOEKETFREKFMOHPKNFLGIASELERKTVACVLYYLTTKKENYKSLVR	484	:	:
Db	433	DRQFMNWTDHEKEIFDKFTQHFPKNFGLIASYLERKSVPDCVLYIYLTKKENNYKALVR	492	:	:
Qy	485	RSY--RRRGSQOQQOQQOQQOQQOQQMPMRSSQEEKDEKEKEKEAEK--BEERPEVEND	541	:	:
Db	493	RNYGRRGRNQIARPSQEEKVEEK-----BEDRAEKTEKKEEKDEEDKDDEDS	544	:	:
Qy	542	KEDLLKKTDDTSEDNDKEKAVASKGRKTSANSQGRKGRITRSWANE--ANSEEAITP	598	:	:
Db	545	KETTCKDRTEATAEPEEREQVTPRGKTKTANSQGRKGRVTRSMTPSAANAATAATE	604	:	:
Qy	599	QQSDEL-----ASMELNESSRWTEEMETAKKGLLEGHRNWSAIARMVGSKTVSOCKNF	652	:	:
Db	605	EPPLPLPPPPISTEPVETSRTWEEMEYAKGLVEHGRNWAIAKMGVTKSEAQCKNF	664	:	:
Qy	653	YFNTRYKRNLDIEILOQHKLMKEERNARRKKKAPAAASEAAPPPVVEDMEASVSG	712	:	:
Db	665	YFNTRYKRNHNLNLQOHKAKSRPREDRYSOCESVASTVSA---QEDBDIAS----	716	:	:
Qy	713	NEEMVYEAEALHASGNVPRGESGPGATVNNSSDTESIPSPh-TEAAK--DTGONGPK	768	:	:
Db	717	NEENPEDSEG-----AENSDDTESAPSPVEAKSESDESSENA--	756	:	:
Qy	769	PPATLGADGPPGPPPTPPRRTSRAPIEPTPASETAGTPPPAPPSPSAPPVYPKKEKE	828	:	:
Db	757	-----ASRGNTPEVAELEATTDPACASP--SSAVPTTKPARRES	794	:	:







Best Local Similarity 35.0%; Pred. No. 8.1e-38;

Matches 401; Conservative 149; Mismatches 393; Indels 202; Gaps 54;

QY 1470 KIDVRSLSGCRTPPPVPHPLVDVAD-ABALERACEY-----ESLKSRCPTASSGGSTA 1523  
DB 7 KINVSLLITGPKSLSGMPLEIVPENIKIVBERGYEDVKAGETVRSRHTSVVSSGSPVL 66  
QY 1524 RGAPVIVPELGRPROSPITYEDHGA-----PEAGHLPRGSPVTMRPTPRLOEGSLSSK 1578  
DB 67 RST---LHEAPKQLSPGIYDTSARRTPVSQNTMSRSPM-----MNRSTDVTISSNK 118  
QY 1579 A-SQDKLSTPRE-----IAKSPHSTVPBHHPHIPSPYHLLRGVSGVDLYRSHIPLAFD 1633  
DB 119 STNHERKSLTPTQRESIPAKSPVPCVDVWVSH--SPFDPHHRGSGTAGVYRSHLPTHL 176  
QY 1634 PPSIGRIPLD-AAAAYLPRHLAPNTVPHLYPVVLINGYDPALEN-ROTIINDYIT 1691  
DB 177 P-AMFPHRALDPAAYLAFQRLSTPGTPGSPQYLY-----AMENTOTIINDYIT 226  
QY 1692 SQOMHNTATAMAQADMLRGLSPRESSIALNYAAGPRGIIDLSOVPHLVLPVPTPGTP 1751  
DB 227 SQOMQVNL-----RPDVARGLSPREQPLGLPYPA-TRGIIDLTNMP-tilvphpggts 278  
QY 1752 ATAMDRLAYLPAQPFSSR-HSSPLSPGCGTHLTKPTTSSSERDRDRDRDRDR 1810  
DB 279 TTPMDRITYPTQITFPFPRPNYSASMSGPHTHL---AAAAASAPERERERERERER 335  
QY 1811 EKSILTTTVEHAPITWRPCTGSSGSSGGSGSSRSPASHASHAHOHSPISPTODA 1870  
DB 336 ---IAAASDL---YLRCPSQO-----PORPOSHGYVRSRSP-SVYTOET 373  
QY 1871 -LOQRPVHLNMGKIITAVPSPKPTVLRSTSPV-----RPAATPPAPCHPLGGLT 1925  
DB 374 MLQORPVSFOGTNGTIVTLPDPTAQLRIMPLAGGSPISQGLPASRYNTAADA-LAALIV 432  
QY 1926 DGYYTLMPEVLLPK-----EAPRVARPERPRAD-----TGH 1957  
DB 433 DAAASAPQMDVSKTESKHEAARLEENLRSAAYVSEQOQLQKTEVEKRSVQCLYTSS 492  
QY 1958 AFLAKPPARGSLPAPASS---PSKSGEPRLVPVPGSHATIARTPAKN--LAPHASPOPP 2012  
DB 493 APFSGKP-----QPHSSVVYSAGDKGP--PPKRYEELTRKTTITANFIDVIT 545  
QY 2013 APPASADPHREKTOSKPFPSIOELRSLGYHGSSYPGVEPVSPVSPSITHDKGLPK 2072  
DB 546 RQIASDKDARERGSSQSSDSS-----SSLSSHRYETPSDAIEVISPASSPAPPQEKLTQY 599  
QY 2073 HLEELDKSHEGELRPQPGPKVLGGEAAHLPHLRPLPESQSPSSPLIQTAPG----- 2125  
DB 600 QPEVVKANAQENDPTQVEGP-----LHRYRP---QOESPSPQOQLPSSQAEQMG 647  
QY 2126 -VKGHVVVTLAHIHSEVITODYTRHH-----PQOLSAPLPAPLYSFGCA--SCPVLDLR 2177  
DB 648 QVPRTHRLITLADHICQITQDFARNQVSSQTPOQ---PPTTFQNSPALSIVSTPV--- 701  
QY 2178 RPPSDLYLPPDP-----HGAP-ARGSPH-----SEGGKRSPEPNKTSVLGGGEDIEPVS 2226  
DB 702 TKTNSRYSPEQAQSVHQRPGSRVSPENLVDKSRGSRPGKSPERSHV---SSEYEPYIS 758  
QY 2227 PPEGTEPGHSRSVAVPILYRDGEOTEPFRMGSKSPGNTSOPPAFPFKLTENSAMVSK 2286  
DB 759 PPQ--VPVVEHKQDLSLLLSQBGAEPAEQNRDARGSPGISYLPSPFTKL-ENTSPMVKSK 815  
QY 2287 KOEINKLNTNHRNPEYNISQPGTEIFENMPAITCTGLMTVRSQAVQEHASTNMGLEALI 2346  
DB 816 KOEIFRKLNSGGGSDMAAAGPGTEIFENLPAVITGSSVGRHGFADPAS--NLGLEDII 874  
QY 2347 RKALMGKYDQWEE-----SPPLSANAENPLNASASLPAAMPITAADGRSDHILTSFGG 2400  
DB 875 RKALMGSDDKVEDHGVMSQPMGV---VPGTANTSV-----VTSGETRREGDPSPHSG 926  
QY 2401 G---KAKVSGRPSRKAAPAP---GLASGDPRPSPSVHSEGDGNCNRRTPLTNRWEDRPS 2456

DB 927 GYCKPKLISNSRKSPIPGQGYLGTERTPSSVSVHSEGDYHRTQP--GWAWEDRPSS 984  
QY 2457 AGSTFPFNPPLMRLOAGVMASPPPPGLPAGSG-PLAGPH---HAWDEPKLLCSQVET 2512  
DB 985 TGSTOFFNPPLTMRM---LSSPTPTPIACAPSAVNOAAPHOONRIWREPAPLLSAQVET 1041  
QY 2513 LSDSE 2517  
DB 1042 LSDSD 1046  
RESULT 4  
T13564  
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N:Alternate names: hypothetical protein EG:49E4.1  
C:Species: Drosophila melanogaster  
C:date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13564  
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: 217689  
A:Accession: T13564  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5327 <SPA>  
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0025392  
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A:Note: EG:49E4.1  
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.3%; Score 574; DB 2; Length 5327;

Best Local Similarity 19.4%; Pred. No. 1.5e-12;

Matches 531; Conservative 332; Mismatches 1123; Indels 752; Gaps 105;

QY 30 IARTHTDVGLL---EYQHSRDYASHLSPGSI-----QPQR 63  
DB 1456 IAKTHKDESSLDKAKEQSRRESLAESIKPESGIDKESALAKESRPSVTDKSEFSR 1515  
QY 64 R-----RP-SLLSEFPQGNERSQELHLRHPESHYSILPELKG 97  
DB 1516 RESTAESLKAESTKDEKSAPPKESRPSGVSVKDETEKSE---PSRRESIAESAK 1571  
QY 98 SEMEPESKRPR--LELLPDLLRP-----SPLA--TGQPAGSEDLTKDRSL--TGKLE 146  
DB 1572 PPIEFREVSREPESVTDGIDKESAKPESRRDSPLASKEASRPESVLESVKDEPIKTEKSR 1631  
QY 147 PVSPSPPTDPELELVPRLSK-----EELION-MDRVDREITMVEQOISKLKKQOOL 200  
DB 1632 RESVAESFKADSTKDEKSPKTSKDISRPESAVENYMDAPFKETSRPESAVGSMK----- 1685  
QY 201 EEEAAKPEPEKPPVPPPIESKHKRSLVQIIYDENKKKAAHRILEGGLGPQVPLYNQ 260  
DB 1686 DESMSKEPSRRESVKDGAQSR-----ETSRPASVAESAKDGADDLKEL---SRP 1732  
QY 261 SDTQYHENIKINQAMRKKLILYFKRNHARKQWKQKFCQYDQLMLEALEKVKVERIENP 320  
DB 1733 ESTTQSKAEGSIKD-----EKSPLEASEAS 1757  
QY 321 RRRAKESVRVYEKQFPIRKORELOERMOSRVQSGLSMSAARSEHEVSEITDGLS 380  
DB 1758 RPASVAESVKDEAKESKEESRESVAEK-----SPLPSKEASRPASVAESIKDEA 1807  
QY 381 EQENLEKQMRQVLAVIPPMYLDADQOQRIKFINMGLMDPMK--VYKQDVNMNMSEQEK 438  
DB 1808 EKSKEESRRESVAESKSP-----PSKEASRPASVAESIKDEAKESKEESRESVAESK 1860  
QY 439 TFRKFMQHPNFGGLIASFLERKTVAECLVYYITTKNENYKSLVRRYSRRRGSKSQOQQ 498  
DB 1861 PLPSKEASRP-----ASVAESI-----KDEAEKS-----K 1885



Qy	499	QOQOQOQOQOQOQPMR\$-----SQEKKDKKEKEKAEKEE---EK---PEVENDKED	544
Db	1886	EESRR\$EVAEK\$PLPSKEASRPASVAESTGKDEAK\$KESRR\$EVAEK\$PLPSKEASRP\$A	1945
Qy	545	LLKEKTD\$TGENDR--KEAVASGRK\$TANSOGRK\$GRI\$TR\$MANEA--N\$EEA\$IT\$QO\$	601
Db	1946	SVAESIKDEAK\$KESRR\$EVAEK\$-PLPSKEASRPASVAESIKDEAK\$KESRR\$E\$V	2004
Qy	602	AELAS\$ELN\$ESSR---WTEEE\$METAK\$GLLEHGRNWSA\$IARMV\$G\$K\$TV\$SOCKN\$FY\$YK\$K	658
Db	2005	AEK\$PLPSKEASRPASVAESIKDEAK\$KESRR\$EVAEK\$PLPSKEASRPASVAESIKD	2064
Qy	659	RQNLDELQ\$H\$K\$M\$KERNAR\$K\$K\$K\$K\$P\$A\$A\$S\$E\$D\$A\$P\$P\$V\$V\$E\$D---E\$E\$M\$A\$S\$V\$G\$N\$E\$E	715
Db	2065	-----EAK\$K\$KESRR\$E\$A\$E\$K\$PL\$P\$--SK\$EASRPASVAESV\$K\$D\$EAD\$K\$E\$E\$SR\$R\$E	2113
Qy	716	EMVEE\$E\$AL\$H\$A\$G\$N\$E\$V\$P\$R\$E\$C\$G\$S\$P\$AT\$V\$N\$N\$S\$D\$T\$E\$S\$I---P\$S\$P\$H\$T\$E\$A\$--K\$D\$T\$G\$O\$N\$G\$P	767
Db	2114	SMA\$E\$K\$A\$Q\$S\$IK\$G\$Q\$S\$P\$L\$K\$E\$V\$R\$P\$S\$V\$A\$E\$S\$K\$D\$D\$P\$V\$K\$E\$F\$R\$S\$R\$E\$S\$V\$A\$G\$T\$A\$D\$A\$R\$D\$Q	2173
Qy	768	K\$P\$AT\$IG\$AD\$G\$P\$P\$P\$G\$P\$T-----P\$P\$R-----R\$T\$S\$R\$A\$P\$IE\$P\$T\$P\$A\$S\$E\$A\$T	803
Db	2174	\$P\$L\$E\$K\$A\$S\$R\$P\$S\$V\$D\$S\$V\$K\$D\$E\$A\$K\$E\$Q\$S\$R\$E\$S\$K\$T\$E\$S\$V\$IP\$P\$K\$A\$K\$D\$D\$K\$P\$K\$E\$V\$LP\$Q\$V\$M\$T\$E\$T\$I	2233
Qy	804	G\$A\$T\$P\$P\$P\$A\$P\$S\$A\$P\$P\$V\$V\$P\$K\$E\$E\$E\$E\$T\$A\$A\$P\$P\$V\$E\$E\$G\$E\$Q\$P\$P\$A\$E\$A\$L\$V\$D\$T\$G\$A\$E\$P\$V\$K	863
Db	2234	\$E\$D\$A\$Q\$P\$M\$P\$S\$Q\$A-----E\$S\$R\$E\$S\$T\$A\$E\$S\$T\$A\$K\$S\$P\$R\$D\$E\$K\$S\$P\$L\$A\$S\$K\$E\$A\$S\$R\$P\$G\$S\$V\$A\$E\$S\$IK	2285
Qy	864	-----S\$E\$C\$T\$E\$A\$B\$E\$G\$P\$A\$K\$D\$E\$A\$E\$A\$T\$A\$E\$G\$A\$K--A\$E\$K\$E\$G\$G\$R\$A\$T\$T\$A\$K\$S\$S\$G\$A	911
Db	2286	\$Y\$D\$L\$K\$P\$Q\$IK\$D\$K\$T\$E\$H\$S\$R\$E\$S\$LE\$D\$K\$S\$A\$V\$T\$E\$K\$S\$V\$S\$R\$P\$L\$S\$V\$A\$S\$D\$H\$E\$A\$V\$A\$E\$D\$A\$K\$S\$S\$IS	2345
Qy	912	P\$Q\$D\$S\$D-----S\$S\$A\$T\$C\$S\$A\$D\$E\$V\$D\$E\$A\$P-----G\$G\$D\$K\$N\$R\$LL\$S\$P\$R\$P\$LL\$T\$P\$T\$G	949
Db	2346	\$P\$K\$D\$S\$R\$P\$G\$F\$V\$E\$T\$V\$S\$S\$P\$I\$E\$A\$T\$M\$E\$E\$S\$K\$I\$E\$V\$E\$K\$S\$L\$A\$L\$S\$Q\$G\$G\$K\$Q\$LT\$D\$S\$P\$V\$D\$V\$A\$E\$G	2405
Qy	950	D-\$P\$R\$A\$N\$A\$S\$Q\$-----K\$P\$L\$D\$L\$K\$Q\$L\$Q\$R\$A\$A\$IP\$T\$Q\$V\$T\$K\$V\$H\$E\$P\$R\$E\$D\$A\$A\$T\$K\$P\$A\$P-----	996
Db	2406	\$F\$H\$A\$V\$A\$S\$V\$T\$V\$T\$P\$T\$K\$P\$A\$E\$A\$Q\$-----I\$G\$A\$A\$K\$T\$V\$S\$S\$P\$D\$E\$A\$L\$T\$P\$S\$A\$P\$E\$H\$S\$R	2455
Qy	997	---P\$A\$P\$P\$P\$Q\$N\$Q\$P\$E\$D\$A\$P\$Q\$G\$S\$S\$P\$R\$K\$S\$R\$P\$A\$P\$D\$K\$E\$A\$F\$A\$E\$A\$O\$K\$P\$G\$D\$P\$C\$M\$T\$S\$G\$L	1053
Db	2456	\$A\$D\$S\$P\$A\$C\$E\$E\$A\$S\$Q\$D\$K\$P\$O\$V\$L\$K\$E\$S\$R---P\$A\$W\$A\$E\$S\$K\$D\$D\$A\$Q\$A\$K\$S\$S\$V\$E\$D-----L	2503
Qy	1054	\$P\$F\$V\$P\$P\$R\$E\$V\$IK-----A\$S\$P\$H\$A\$D\$P\$D\$S\$A\$F\$Y\$A\$P\$P\$G\$H\$P\$L\$P\$L\$G\$L\$H\$D\$T\$A\$R\$P\$V\$L\$P-----R	1098
Db	2504	\$R\$P\$V\$A\$S\$T\$E\$I\$S\$P\$A\$S\$A\$G\$E\$T\$A\$S\$P\$E\$E\$A\$P\$K\$D\$F\$A-----E\$F\$E\$Q\$A\$E\$K\$A\$V\$L\$P\$L\$T\$I\$E\$L\$K\$G\$N	2553
Qy	1099	\$P\$T\$I\$S\$N\$P\$P\$P\$L\$I\$S\$A\$K\$H\$P\$S\$V\$L\$E\$R\$Q\$I\$G\$A\$I\$S\$O\$G\$H\$S\$V\$Q\$L\$H\$V\$P\$Y\$S\$E\$H\$A\$K\$A\$P\$G\$V\$P---T\$M\$G\$L\$P\$L\$P	1155
Db	2554	\$L\$P\$T\$L\$S\$S\$P\$V\$D\$V\$A\$H\$A\$S\$V\$Q\$P\$A\$E\$L\$K-----V\$D\$I\$E\$K\$T\$A\$S\$P\$D\$E\$A\$K\$S\$L\$IG\$S\$P\$A\$E	2599
Qy	1156	\$M\$D\$P\$K\$L\$A\$P\$G\$V\$K\$Q\$E\$L\$S\$P\$R\$G\$A\$G\$P\$E\$S\$IG\$V\$P\$T\$A\$Q\$E\$A-----S\$V\$L\$G\$T\$A\$L\$G\$S\$V\$P\$G	1205
Db	2600	\$E\$R\$P\$E\$S\$A\$E\$S\$A\$K\$D\$A\$E\$S\$V\$E\$K\$S\$D\$A\$S\$R\$P\$S\$V\$E\$S\$T\$A\$D\$S\$T\$K\$G\$D\$I\$S\$P\$S\$E\$S\$V\$E\$G-----P\$K	2653
Qy	1206	\$G\$S\$T\$K\$IG\$P\$T\$R\$V\$P\$S\$D\$A\$I\$Y\$R\$G\$S\$I\$T\$H\$G\$T\$P\$A\$D\$V\$L\$K\$G\$T\$T\$R\$IG\$E\$D\$S\$P\$R\$D\$R-----	1257
Db	2654	\$D\$V\$E\$K\$S\$E\$S\$R\$P\$P\$S\$V\$S\$A-----S\$I\$T\$G\$D\$T\$K\$D\$V\$S\$R\$P\$A\$S\$V\$V\$E\$S\$K\$D\$E\$H\$D\$K\$A\$E\$S\$R\$R\$E\$S\$T\$A\$K\$V\$E	2708
Qy	1258	-----G\$R\$E\$D\$S\$L\$P\$K\$G\$H\$V\$Y\$E\$G\$K\$H\$V\$L\$S\$E\$G\$M---S\$V\$T\$Q\$C\$S\$K\$E\$D\$G\$R\$S\$S\$G\$P\$P\$H\$E\$T\$A\$P\$K\$R	1309
Db	2709	\$S\$V\$D\$E\$A\$G\$K\$D\$S\$K\$S\$S\$O\$Q\$D\$K\$E\$S\$T\$Y\$A\$K\$E\$A\$S\$R\$R\$E\$S\$V\$V\$E\$S\$S\$K\$D\$D\$A\$E\$K\$S\$R\$E\$S\$P\$E\$S\$V\$H\$A---	2765
Qy	1310	\$T\$Y\$D\$W\$E\$G\$R\$V\$G\$R\$A\$I\$S\$S\$A\$S\$T\$E\$B\$G\$L\$M\$G\$R\$A\$I\$P\$P\$E\$R\$H\$S\$P\$H\$H\$K\$E\$Q\$H\$IR\$G\$S\$IT\$Q\$IG\$P\$R\$S\$V\$E\$A\$Q\$E	1369
Db	2766	-----S\$G\$E\$P\$V\$P\$R\$E\$S\$K\$S\$P\$L\$D\$S\$K\$D\$T\$S\$R--P\$C\$S\$V\$V\$E\$S\$V--T\$A\$E\$D\$E\$K\$S\$E\$Q	2802
Qy	1370	\$Y\$L\$R\$E\$A\$K\$LL\$R\$G\$T\$P\$P\$P\$P\$P\$S\$D\$L\$T\$E\$A\$Y\$K\$T\$Q\$A\$L\$G\$P\$L\$K\$P\$A\$H\$E\$G\$L\$V\$A\$T\$V\$E\$A\$G\$R\$S\$I\$H\$E\$I\$P	1429
Db	2803	\$Q\$S\$R\$E\$S-----V\$A\$E\$S\$V\$K\$A\$D\$T\$K\$D\$K\$G\$S-----	2823

Qy	1430	REELRHTPELPLAPRPLKGGSTQGTPLKYDT-GASTTGSKKHDVRSGLSGPGRFTFPVH	1488
Qy	1431	REELRHTPELPLAPRPLKGGSTQGTPLKYDT-GASTTGSKKHDVRSGLSGPGRFTFPVH	1489
Db	2824	-QE-----ASRPSSYDELLKDDDEKQESRRQISITGSHK-----	2855
Qy	1489	PLQVMDARALERACYEESLKRPGTPASSSGGSIGARGPVTVPELKGKPROSPL--TYEDH	1546
Db	2856	AMSTMGDSPMDKADSKKE-PGRPSVAES-----IKHENTKDEESPLGSRDVS	2904
Qy	1547	GAPFAGHLPRG--SPYTMREPT-PRLQEGSSSSKASQDRK--LTSTPRELAKSPHSTVP	1601
Db	2905	AESIKSDITKGEKSPPLPSKEVSRPESVWGSIKDEKAESRRRESVAESVKPESSKDATSAPP	2964
Qy	1602	-EHHPHIPISYEH-----LRGVSGVDLYRSHIPL-----AFDPPTSIPRGILPDAAA	1647
Db	2965	SKEHSRPESVGLSKDEGDKTTSRRYSVADSIKDEKSLLVQEAASRPSEAESLK--DAAA	3023
Qy	1648	AYYLPHILAPNPYTHLYPYLLIRGYPTAALEENQTIINDYIISQOHHHTATAMAQRA	1707
Db	3024	-----PQETSRPESVTESVKDGKSPVAKESRASPASVAENAKOSADESKSP	3072
Qy	1708	DMRLGLSPRESSIALNAAAGPRGIID-----LSQVPHLPVLVPPYPTGTPAT	1753
Db	3073	ESL-----POSKAGSINKDEKSPASKDEAEKSKESRRRESVAEQFP---LYSKEVSRPAS	3124
Qy	1754	AMDBL-----AYLTPAPOFFSRHSSSPSPGPGTHLTKTPTTSSSERERDRDRDRDR	1810
Db	3125	VAESVKDEAEKSKESPLMSKESKASPASVAG-----SVKDEAEKSKESR-RESVA	3174
Qy	1811	EKSLTITTTTVEHAPITWRPGTQSSSGSSSGGGGSSRRPASHISHAHQHSPISPRTQDA	1870
Db	3175	EKSPPLSKESRASPASVAESVKDEAKSKESRRRESGAEKSPASKESRASPASVAESIKDE	3234
Qy	1871	LQ--QRPSVLHNTGMGIITAVEPKSPVL-----RSTSTSSPVVRPAATPPATHC	1919
Db	3235	AEKSKESRRRESVAESKPLPSKESRPTSVAESVKDEAEKSKESRRSDSVAEKSP-----	3289
Qy	1920	PLGGTLGVVYPTLMPEVLLPKCAPRVARPERPRADTGHAFIAPKARGGLEPASSP---S	1976
Db	3290	-----LASKESRASPASVAESVQD--EAEKSKESRRRESVAESKSPLAYK	3330
Qy	1977	KGSEPRPLVPVSGHATARTAKNLAPHIASPDPPAPPASADPHREKTKQSKFSTOEL	2036
Db	3331	EASRPASVAESIKDEAEKSKESRR-----ESVAEKSPASKESRPTSVAES	3378
Qy	2037	ELRSLGVHGSSYSPGEPVPSVSPSPSLTHDKLPKHL--EELDKSHLEG-----ELRKPQ	2092
Db	3379	VKDEAEKSKESRRSDSVAESKPLASKESRASPASVAESVQDEAEKSKESRRRESVAESPL	3438
Qy	2093	PVKLGGAHLPHL-----RPLPESQSPSPLLQTPAGVKGHQVVTLAQHTISEV	2142
Db	3439	ASKEASRPASVAESVKDDAEKSKESRRRESVAESKSPAS-----KEASRPASVAESVKDE	3493
Qy	2143	I--TODYTRHHPOOLASPLPAPLYSFPAGSCVLDLRPPSDLYLPPDHGAPARGSPHS	2200
Db	3494	AEKSKESRRRESVAESKSPPLPSK-----EASRPTS-----VAESVKD	3529
Qy	2201	EGGKRSPENKTSVLGGEDGTPEVSPBGMTEPGHRSRASYVLLYRDGEOTEPSRMGSK	2260
Db	3530	EAEKSKESRRRESVAESKSLASKESRASPASVAE-----SVKDEAEKSKESRRRESVAEK	3583
Qy	2261	SPGNTGPPAFFSKLTESNAMYKKOEINKKLNTHNRNE-----PEYNISQPGTEI	2313
Db	3584	SP-----LASKESRASPASVAESVKDEAEKSKESVSRRESVAESKPLPSKEASRP----	3631
Qy	2314	FNMPAITGTLWYRQAOVEHASTNMGLEAIIRKALMGKYDQWEEESPPLSANAFNPLNA	2373
Db	3632	-----TSVAESVKDEADSK-----EESRRRESGAEKSP--	3660
Qy	2374	SASLPAAMPITAADGRSDYHTLTPSGGGGKAKYSGRPSRKAASPAAGLASGRDPPSPS-S	2432
Db	3661	-ASMEASRPTSVAESVKDET-----EKSKEESRRRESVTEKSPUPS--KEASRPTSVAES	3711
Qy	2433	VHSEGCNRR-----RTPLNRYWEDRPSSAGST	2460











QY 1685 IINDYITTSQMHNTATANAQADMLRLGLSPRESSIAL---NYAAGPRGIIDLSQVPHLP 1741  
Db 1214 GSNVGMEDGERPRRRHGRAQODK---PPFRRLKQERENARGSEG----- 1258  
QY 1742 VLVPTPTGTPATA---MDRLAYLPTAPQPFSSRHSSSPSPGPPHLTKPTTSS----- 1793  
Db 1259 ---KPSLTLPASAPGPEEALTITVAPAPRAAAKSPDLNSQNSQANQEWETASESSDF 1315  
QY 1794 -SERERDRDRE-----RDRERERKSILTSITT 1820  
Db 1316 TSERRGDKKAPPVLLTPKAVGTGGGGGAVPGTISAMSRGDLQRAKDLKRSFSSQRP 1375  
QY 1821 VEHAPITWRPGTBSGSSSSSGGGG-----SSSRPASHASHAQ 1859  
Db 1376 GMERQNRERPGGKAGSGSSSGGGGGPGGTGRCGDKRWPSPKNSRPERPERPGL 1435  
QY 1860 HSPISPTODALQORPSVLHNTGMKIITAVBPSKPTVLRSTSTSPVNRPAATFFPATHC 1919  
Db 1436 PLPPPPPPSSSAVRLDQVTH-----SNPAGIQALQALSSRQGSVTAPGGH- 1481  
QY 1920 PLGGTLDGYPTLMEPVLLPKAPVARPERPRADTGHAFKAPPAR-----SGLEPA 1972  
Db 1482 -----PR-HKPGPPQAPQGS--PRPTRYEPQRVNSGL--- 1512  
QY 1973 SSPSKGSEPRPLVPVSG-----HATARTPAK-----NLAPHAS----- 2008  
Db 1513 SSDPHEEPGMVGVGCTPRDSAGVSPFPKRRRPPKPELLQESLPPPHSSGFLGS 1572  
QY 2009 -PDPAPPASAD-----PH-----REKTSKPFISQ-----ELE 2037  
Db 1573 KPEGPGQAESRDTGTEALTPHWNRLHATSRKYSRYPTSMEMPWMEPLSPFFEDVAGTEMS 1632  
QY 2038 LRSGLYHGSYSPEGVSPVSSSLTHDKGL-----PKH-----LEELDKSHLEG 2084  
Db 1633 QSDSGVDLSDGSQVSSGPGCSQRSSP-----DGLKGAAEGPPKRRGSSPLNAVPCGGPG 1688  
QY 2085 ELRPQKPGVKLGCGAAHLPHLRPLPESQSPSSPLLQATAPGVKGHVTLAQHISEVIT 2144  
Db 1689 SEPPRRPPAPHDGRKELPRQPLP-----PGPIGTER---SOR-----T 1726  
QY 2145 QYTRHHPOOLSAPLAPLYSPFGACPVLDLR-----PPS-DL 2183  
Db 1727 DRGTEPGPLRPS-HRPGPPVQ-GTSDKSDLRVLVGDLSKAEKELTASVTEAIPVSRDW 1784  
QY 2184 YLPPPDHGAARGSPHS---EGGKRSRPN-----KTSVLGGGEDGI- 2222  
Db 1785 ELLP---SAAASAEFQSKNLDGCHCVPEPSSGQRLYPVFTVGSAGPSSQISGSGHGLS 1841  
QY 2223 -----EPVSPPEGTEPGHRSRAV-YPLLRYDGEQTEPFSRMGSKS 2261  
Db 1842 ITSQWRLRPGTSLHYPYRSQPLYLPPGPAPPSALLSGVALKGQFLDFSTMQATELGKLP 1901  
QY 2262 PONTSOPPAFF-----SKTESNSAMVSKKQEIINKLTHNRNEPEYNISOPGREI 2313  
Db 1902 AGGVLYPPPSFSLYSPAFCSPLPDTSLQVRO-----DLPSP-SDF 1941  
QY 2314 FNMPAITG--TCLMTRYSOAVOEHAETNMGLEAIIKALMGKYDQWESPELPSANAFNPL 2371  
Db 1942 YSTPLQPGQSGFLSGAPAQO-----MLLPWDSQLPVNFGSLPPAPP---PAPPL 1992  
QY 2372 NASASLPAAMPITAADGRDHTLTSPGGGKAKVSGRPSRKAAPGLAGDGRPPSVS 2431  
Db 1993 SLLPVGPALQPPSLA-----VRPPPAATRVLPSPA-----RPFPA 2029  
QY 2432 SVHSE-----GDCNRRTPLNRVWEDRPPSAGSTPPFPYNPLI--M 2469  
Db 2030 LGRAELHVPVLPKFDQYKLSNLGGPSSRTPPTGRFSGLNSRLKATPSTYSGVFRTO 2089  
QY 2470 RLQAGVMASP-----PPPLPAGSGPLAGPHHAWDEP 2502  
Db 2090 RVDLYQOASPPDALRWPKPWERTGPPREGFSRR-AEEP 2128

RESULT 7  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N:Alternate names: alpha-NAC protein  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
C:Accession: T30826  
R:Yotov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a mus  
A:Reference number: Z20889; MUID:96312450  
A:Accession: T30826  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2187 <YOT>  
A:Cross-references: EMBL:U48363; NID:gl666688; PID:gl666689; PIDN:AA18732.1  
C:Genetics:  
A:Gene: Naca  
A:Map position: 10  
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding ac  
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 3.9%; Score 520; DB 2; Length 2187;  
Best Local Similarity 19.7%; Pred. No. 4.4e-11;  
Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;

QY 738 GPATVNNSDTESIPSPHTEAA-----KDTGQGPKPATLGADGP---PGPPT 784  
Db 3 GEATETVPATEQELPQOAEATVLPMSALKVAAVGQGPPTPPSSILGPOQSPIVTAHQPS 62  
QY 785 P-PRRTSRAPIE-----TTPASEA---TGATTP-----PP----- 810  
Db 63 PLPSSVSTPPEVPPAQPIATAETALPSTGATPTPTFLPHLIGPPISPAALALASPMIGLA 122  
QY 811 -----APPSAPPPVVPKKEEETAAAPVE----- 838  
Db 123 QKGARSSAPLSLVALAPHSVQKSSVCPHPPLTSPSAAAGELGALTASIPLEPKTSTS 182  
QY 839 -----EGEEKQPP-----AAEELAVDTGKAEPEPVKSE 865  
Db 183 QVPSQGTNLNLAGTAPCPDVVVRAPPSHLENPLASVQPGMLSCPQTLNTPSVKGVPISSA 242  
QY 866 CTEEAEEGPAKGDAAEAATAEGAKAE-----KKEGSGGRAT 904  
Db 243 LTQSRLSLNLKGPVSPPARNTAAPSIIAPLSTSLGCHLPLLHHSSVDSPPIPPGSGGLAV 302  
QY 905 T-----AKSSGAPQSDSDSSATCSADEVDEAEGDK- 934  
Db 303 SNPTSVGHSGIAASCPPERCVCVVPALPSRLLAVDSCAAPDDKGSASVTNELCSPPGSSNV 362  
QY 935 -NRLLSRPRLTPTGDPNANASPOKPLDLKOLKORAAAIPPIQVTKVHEPP----- 985  
Db 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASOKTGLKEIPVSCIGATH 409  
QY 986 ---REDAAPTKAPPAPPPQNLQPEDAPQPGSSPRGKSRSPAPPADKAEFAAEAKQ 1041  
Db 410 HALDNPSAISVAPATHVPPPT-----SSGLVSKDPSASPTSLVVPAAHKQ 455  
QY 1042 LPGDPCCWTSLGPF-PVPPREVIKASHAPDPSAFSPAPPGHPLPLGL-----HDTARVPL 1096  
Db 456 FPAPPASATLGVVPSPLPATEGLK---NLPISALVNVGAPVSPAQAAGLPTKRKDTTLQPLA 512  
QY 1097 P-----RPTIISNPPPLISAKHPSVLERIGIATISQMSVOL 1133  
Db 513 PIALKESPSOSASLSLEVLSEDTVTTKTGGPAPVVRPA-----IAGVATTTSURA 563  
QY 1134 HVPYSE-HAKAPVGVMTGLPL-----PMDPKKLAPFSGVKQEQLSRPGQAGPPE 1182  
Db 564 DSPPAIVRADSCVSNPTVSQPLKRSVTDPAAMAPRTAKNTAPST---TSPLVPLASEGCPV 620  
QY 1183 SLGVPTAQEASVLRGTALGVSPPGSGITKGP-----STRVPSDAITYRGS 1229



Db 621 ASSMALSPQNASVSEFATALALSP--EIPKSVPPDPPLAEISFSNARKYDAVSHMESSGS 678  
QY 1230 THGTP-ADVLYKGTTRIIGEDSPSLDRGREDs-----LPKGVHI 1269  
Db 679 ROGHDPASVTAGTVVCL-----ADSLDTSVSKGSGALSASSPLYPLEVSFLPEAGLA 734  
QY 1270 YGKKGHVLSEGGMSVTOCSKEDG--RSSSGPPHETAA----- 1306  
Db 735 VOGPKGSL-----NKLSPTPSSKGAPVPSTGAPPSPKGAPIVPTESSISKQVPAEILPS 790  
QY 1307 PKRTVDMGGRVGRATISSAIEGLM-----GRAIPERHSPHILKEQHHI--- 1351  
Db 791 POKTEVTASRLISAVQSPKVDPIMSDVTPTSPKKTATAPKDTSATLSLKSVPVAVTSL 850  
QY 1352 -----RGSITOGIPR----- 1361  
Db 851 SPPKAPVAPSNEATIVTPTISLKNALAAATPKETLATISIPKVTSPSPQKTPKSVSLKG 910  
QY 1362 -----SYVEAQED-----YLREAKLLKRE-----GTPPP--pp 1388  
Db 911 APAMTSKKATEIAAASKDVSPSPKFPKPEVLLQHVPTSPPKSPVSDTLGALTSPPPKGGP 970  
QY 1389 PSRDLTEAVKTOALGPKLK-----PAHEGLVATYKEAGRSIETPR-----EELRHTPEL 1439  
Db 971 ATLAETPTTPKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKTAAPKRESSATSS 1030  
QY 1440 PLAPRLKBSGSIQ-----GTPLKYDTGASTGSKKHVRSLSIGSPGRT----- 1483  
Db 1031 KRAPKAVTSKEIPSKGVTAVPLEISLPLKETSKSATPGEKS-----ASSPKRSPTAGP 1084  
QY 1484 --FPP-----VHPLDMADARALERACVEESL-----KSRPGTA-----SSSGGSJAR--- 1524  
Db 1085 KETPPGGVTAVPPEISLPPKETPQONATPNESLAASQKRSKTPVPKETPPGGVTAMPLE 1144  
QY 1525 --GAPVIVPELCKPROSPITYEDHAGPAGHLPGRSPVTMREPTPRLOEGSLSSKASOD 1582  
Db 1145 IPSAQOKAPKTAVPKOIP--TPED-----AVTLAGSLSPSPK-----ASKTAAP 1187  
QY 1583 RKLSTTPR-----ETAKSPHST-----VPEHHPHPISPYEHLRLRGVGV 1621  
Db 1188 KEAPATPSGVTVAVSGEISPSPKTSKTAAPKENSATLPPKSPKTAAPKETAPTSSEGV 1247  
QY 1622 DLYRSHIPAFDPTIPRIGIPIDAAAAYLPHRLAPNPTYPHLYPPYLRIGYDPTAALEN 1681  
Db 1248 TAVPSEISPS--PPTASKGVV--TLTPKGAPNALAESPAASP-----KKVPKTAAPPE 1297  
QY 1682 ROTIINDYTSQMHNTATAMAQADMLRGLSPRESSLALNYAAGPGRIIDLSQVPHLP 1741  
Db 1298 TST-----TPSPQKIPKVG--PKEAS----- 1317  
QY 1742 VLVPPPTPGTPATMDRLAYLPT-----APOPFSSRRHSSPSLPGGPTHLTPTTT 1791  
Db 1318 -ATPSSKTPKTAVKETSAPESGVTAVPLEIPPSPRKAPKTAAP-----KETPA 1366  
QY 1792 SSSERDRDRDRDREREKSILTSTTVEHAPITWRPGTEQSSGSSGGGGSSSRP 1851  
Db 1367 PSEGE-----ATTAPVQIPPSPRKSGKA-----GSKETP 1396  
QY 1852 ASHS-----HAHQHSPISPRTODALQORPSVLHNTGMKGIITAVPSPKTVLRSTSSP 1906  
Db 1397 TTPSPGVTAAPLEIPISSKKTSKMASPKETLVTPSSKKLSQVTPKETSLEGATAVPLE 1456  
QY 1907 VSPAATFPPAT-----HCLPGGLTLDGYVPTLMEPVLLPKEAPRVARPERPRADTGHAFK 1962  
Db 1457 IPPSHKKAPKTVDPKQVPTSPKDAPTTLAESPSPPKAPKTAAPPSERVTT-----V 1510  
QY 1963 PPARSGLPASPSSKSGSEPRPLVPVSGHATITARTPAKNLAPHASPDPPAPPA--SASD 2020  
Db 1511 PP-----EKPATPKASGTASKVPV-----PAETQEVAVSRETPTVPVPPVKN 1556  
QY 2021 PHREKTSQKPSIQLELRSLGYHGSSYSPEGVEVPSVSPSLTHDKLPLKHELDKS 2080

Db 1557 PSSHKTKSTKIELKE-----APATLPP--SPTKSPKIPSSKKAPR----- 1594  
QY 2081 HLEGELRPKQPGVPLGGEEAHLPHLRPLPES--OPSSSPLLOTAPGVKGHQRVVTLAQHI 2139  
Db 1595 -----TSAPK-----EFPASPSIKPVTTSLAQTAPPSLQKAPS----- 1627  
QY 2140 SEVITQDYTRHHPOQLSAPLAPLPLYS--FPGA-----SCP 2172  
Db 1628 -----TTIPKENLAAPAVLPVSKSPAPARASASLSPATAAPOTAPKATTIPSPCK 1679  
QY 2173 VLDLRRPPSDLYLPDPDHGAPARGSPHS-----EGGKBS-----PEPNK 2211  
Db 1680 KAAATETPIETATAPSELEGAPKETSETSVKVLMSPPPKASSSKRSTASTPATLPSLKE 1739  
QY 2212 TSVL-----GGGDEGIEPVP--PEGTEPGHRSVAVPLLYRDGEOTEPSRMGSKSPG 2263  
Db 1740 ASVLSPTATSSGKDSHISVSDACSTGTTTPQASEKL-----PSKKG----- 1781  
QY 2264 NTSQPPAFFSKUT--ESNSAMVSKKQEIKNKLTHNRNEPEYNIQSQTEIFNMPAITG 2321  
Db 1782 ----PTAFTTEMLAAPAPESALAITAPIQKSPGANSNSASSPKCPDPSSKKDTKGLPS--- 1834  
QY 2322 TGLMITYRSQAVQEHASTNMGLEAIRKALMGKYDQWEESSPPLSANAFNLNA----- 2373  
Db 1835 --AVALAPOTVPEKDTSKAIETLLVSPAKGSDCLHSPKGPVGSQVATPLAFTSKVPP 1892  
QY 2374 ---SASLPAAMPITAADGRSDHTLT---SPGGGKAKVSGRPSRRKAKSPAGSLASGDRP 2427  
Db 1893 EAVSASV-APKAPAA-----SLTLAPSP-----VAPLPKQPLLESAPG--SVLES 1936  
QY 2428 PSVSSVHSGEGDCNRRPTPLNRVWEDRPPSAGSTPPFPYPLNLRQLQAGVNASPPPGGLPA 2486  
Db 1937 PSKLPVPAEED--ELPPLI-----PPEAVSGGEPFQPILVN-----MPAPKPACTPA 1981

RESULT 8  
A1819  
proline-rich peptides 637K precursor, prostatic - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-May-1992 #sequence revision 13-May-1992 #text\_change 05-Nov-1999  
C:Accession: A42663; A41819; A31966; B20593; A20593  
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.  
J. Biol. Chem. 267, 9884-9894, 1992  
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are  
le exon.  
A:Reference number: A42663; MUID:92250652  
A:Accession: A42663  
A:Molecule type: DNA; mRNA  
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>  
A:Cross-references: GB:M86514  
A:Experimental source: ventral prostate  
A:Note: sequence inconsistent with the nucleotide translation  
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.  
Submitted to the Protein Sequence Database, April 1992  
A:Reference number: A41819  
A:Accession: A41819  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-5762 <DEL>  
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; W  
J. Biol. Chem. 263, 19159-19165, 1988  
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich  
A:Reference number: A31966; MUID:89066721  
A:Accession: A31966  
A:Molecule type: mRNA  
A:Residues: 3372-3540 <HEM>  
A:Cross-references: GB:M20721; GB:J04188; NID:g206397; PIDN:AAA41950.1; PID:g554494  
R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.  
J. Biol. Chem. 258, 14206-14211, 1983  
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primar  
A:Reference number: A94675; MUID:84061859  
A:Accession: B20593























D	b	6264	KKEVPPPKVPPEEPKPVPEKKVPVKVIMKEBPLPAKVTEKHMQITQBEKVLVAVTKKEA	63232
Q	y	861	PVKSECTEE-----AEEGPAKGDAEAAEATAEG-ALKAEEKEGSGR	902
D	b	6324	PPRARVPEEPKRAVPEEKVLKLKPREEEPAPKVTFRKRVVKEEKYSIEAPKEPQPIK	6383
Q	y	903	ATT'---AKSGAPQDSDSATCSADVEDBAEGDKNRLLSPRPSLLTPTGDPRAN----	954
D	b	6384	EVTIMEEKERAYTLEEAHSVOREEYEBEYDYKEFEYEP---THEVDQYEYEBERE	6440
Q	y	955	-----ASPOQLDLKOLKORAA-----AIP-----PIQTWIKVEHP	985
D	b	6441	YERYEEHEVITEPEKAPIVKVPPEEPVTPKAPPKVLKKAIVPEKVPVIPKKLUKP	6500
Q	y	986	-----REDAA--PTKP-----APP-----	997
D	b	6501	PPKVPEEPKKVPEEKITHISITKREEQVTEPAKVPMPKRVAEEKVPVPRKEVAPVR	6560
Q	y	998	APPPQNMLQESDA-----POQG	1016
D	b	6561	VPEVPKELEPEEVAFEEFVVTHVEEYLVEEBEETHIEEETIIEEVVVPVPRKEVVP-	6619
Q	y	1017	SSPRGKSRSAP-PADKEFAAEAKLDCDP-----PCWT	1050
D	b	6620	RKVPEEKKVPVPPKKKEAPPKAVEPVRKKPEKVPVLI PKKEKPPPAKVPEVPKKVPE	6679
Q	y	1051	SGLPFPVPPREVIAKSP-----HAPDPSAFSYAPGGH--PLPLGLHDTA	1092
D	b	6680	EKVPVVPVKK--VEAPPKAVEPVPKKVPPEKKVPVPAKKVE-APPKAVEPVPKKLPEE	6736
Q	y	1093	RPVLPRPPTISNPPPLISSAKHPSVLEROIGAISOGMSVOLHVYPYSEHA--KAPVGPTM	1150
D	b	6737	KKPTVPVKVEAAPPPKVPKKREP-----VVPVALQOEVELFVEEIVPEEE	6783
Q	y	1151	GLP-----LPMDDPKLAIFSGV--KOELSPRGAGPESLGSVTAQEASVLRTALGSV	1203
D	b	6784	VLDEEEVLPEEEVLPEEEVLPEEEVLPEEEVLPEEEVLPEEEVLPEEEVLPEEEVLPEEE	6833
Q	y	1204	PGGSITKGI--PSTRVPS-----DSAITYRGSI-----THGTADVLTKGTITRIIGEDSPS	1253
D	b	6834	PEEEVLPEVKVPVVPAPVPAPEIKKKVTEKKVVI PKKEAPPAKV-----PEVPK	6882
Q	y	1254	RLDRGEDSLPKGHVIYEGKGHVLSYEGGMSVTOCSEDGRSSGGSPPHETAAPKRYTDM	1313
D	b	6883	KVEEK-----ILPKEEVELPVE---VTEPEEPEISEEIEPEEPSTSEEVEE	6929
Q	y	1314	MEGRVGAIRASSIEGLMGRAJPPRHSPHHLKQHHRIGSIGTIDPSYVVEAQEDLYRR	1373
D	b	6930	APRVPVEIKKAVEA--PTPVPKVEAP-----PAKVSKKIPEKVPVVPVQ--KK	6976
Q	y	1374	EALLKREGTTPPPPSRDLTEAYKTQALGPLLKPAAHEGLVA-----TVKE-----	1420
D	b	6977	EAPPKAVEPVPKKVPEKVLVP--KKEAVPPAKGTVLVEKVSVAFRQEVVVVERLEEV	7034
Q	y	1421	AGRSIHIEIPRELHRHPELPLAPRLPKBSGITQGTPLKYDTGASTTGSKHKHDSRLGISP	1480
D	b	7035	VEAEVEIPEEEREFHEVE-----EYFEGEF-----HEVEEFIKLE	7070
Q	y	1481	GRTTFPPVHLDD---VMADARALEACRYEESLSRPGTAGSSGGSIARGAP--VIVPELG	1534
D	b	7071	QHRVEEHREKVCVHRIEIVEFAEEVEVEFKP-KAPPKGEISEKIIPP KPTPKPRKE	7129
Q	y	1535	KPROSP-----LTVEDHCAPFAGHLPGSPVYTMREPTPLQEGSLSSSKASQDRKLTS	1588
D	b	7130	PPAKVPEVPKKI VEEKVRVP--BEPRVPPTKVPVVLPPKVEVPEKKVPVPAKPRPEAP	7186
Q	y	1589	PREIAKSHSTVPEHHPHIPISPYEHLLRGVSGVDLYRSHIPAFDPTISPRGIPLDAAA	1648
D	b	7187	PPKVPEAPVEVPKKV-PVPPPK-----KPEVP----PTKVPE-VPKAAVPE	7228
Q	y	1649	YYLPRHLANPTYPHYLPYPLYRGTIRGYPDTAALENRQTIINDYITSQQMHNTATAMAQRAD	1708
D	b	7229	KKYPEAI PPKPESP---PPEVEF-----	7248

QY	1709	MLRGLSPRESSLA-----LNYAAGPRGIIDLSQVPH-----LPVLVP-----	1745
DB	7249	-----EPEESPSAPPKPEPPVVRVPEVKVPEKPPAAAPKKPEVTPVKVPEAPKEV	7303
QY	1746	-----PTPGNPATAMDRLAYLPTAPOPFSSRHSSSPLSGGTHLTTKTTTTSSERE	1797
DB	7304	VPEKKVPVPPKPEVPPPTTKVPEVKVAVPEKKVPAIDPKPESP-----PPEVPEEPPEV	7359
QY	1798	RDRDRDRDREREKESILSTTTTTHAPITWRPGTEQSSGSSSGGGGSSSRPASHSHA	1857
DB	7360	ALEPPAEVVEEPEAPPAQVTPPKNPV-----PEKKAPAVVAKK	7400
QY	1858	HQHSFISPRQDALQORPSVLHNTGHMKGIITAVERSKP-----TVLSTSTSS	1905
DB	7401	PELPFV-----KVPEVKVPEVKVPLVPKKKPEAPPAKVPEVKVPEKKVAV	7451
QY	1906	PVRPAATFPPTHCPGLGTTLDGVYPTLMPEPVLPLPEAPVARPERPRADTGHAFLLAKPPA	1965
DB	7452	PKKPEV--PPAK-----VPEVPKKPVL--EEKPAVPPEAESEPPPEVY--EEPE	7495
QY	1966	RSGLPEASSPSK-----GSEPRPLVPVPSGHATIAITPAKNL-----APHHA	2007
DB	7496	ETAPEEETAPEEKPVVAEEEPPEVPP---PAVPEEPKIIPEKKVPVKKPEAPPPK	7551
QY	2008	SPDP-----PAPPASADSPHEKTSQKPSFI-----QELLSRLG	2042
DB	7552	EPEEKVIEKKLKPURPPPPPPAPPKEDVKKEIFQLKAIPKKKVPENQVPEKVELTPLK	7611
QY	2043	YHGS-----SYSPGVEPVSPVS---SPSLTHDKGLPKHLEELDOKSHLGEELRPK	2089
DB	7612	VPGGKKVKRLPERKEPEKPEEVVLSVLKRPEEEEPKVEPKLEKVKKPAVP---EPP	7668
QY	2090	QPGPVK-----LGGEAAHLPHLRPLPESQSSPLLQTAPOGVGHQRVVTLAGHISEVI	2143
DB	7669	PKKPEVEEVPTVTKRKIKPEPKIPEKIPK--AIPLPAPPEPKPEAEVKT-----I	7719
QY	2144	TQDTRHHQOOLSAPLPAPLYSPFG--ASCPVLDLRRPSSDLYLPDPDHGAPARGSPHSEG	2202
DB	7720	KPPVPEPPIAPVTVPVVGGKAKAKAPEAAKPK-----GPIKGV-----	7764
QY	2203	GKRSPENKTS--VLGGGEDGIEP-----	2224
DB	7765	-KKTSPSTEARRKLKRGSGGKPPDEAPPTYQLKAVPLKFVKIKEIDIIITSESEFVGSSA	7823
QY	2225	-----VSPPECMT-----EPGHSRSA-----VYPLYLRDGGQTPESRMGS	2259
DB	7824	IFECLVSPSTAITWMKDGNIRESPKHFRFADGDKRKLHIIDVOLSDAGEYTCVLRIGN	7883
QY	2260	KSPGNTS-----QPPAFPSKLTESNAMYKSK---KOEINKKLTHNRNPEYNISQPG	2310
DB	7884	KEKSTAKLVVEELPVRFKVTLSEEVTVKGGQPLYLSCELNNKRDVVRWKDGKIVVEKPG	7943
QY	2311	TEIFNMPAITG 2321	
DB	7944	RIV--PGVIG 7951	
RESULT 12			
S37671			
MHC class III histocompatibility antigen HLA-B-associated protein 2 [simila			
C:Species: Homo sapiens (man)			
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000			
R:Bougueleret, L.			
submitted to the EMBL Data Library, August 1992			
A:Reference number: S37671			
A:Accession: S37671			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1870 <BOU>			
A:Cross-references: EMBL:Z15025; NID:G29374; PID:G29375			
C:Genetics:			



A:Map position: 6p21.3  
A:introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65  
C:Superfamily: collagen alpha 1(IV) chain

Query Match 3.8%; Score 503; DB 2; Length 1870;  
Best Local Similarity 21.2%; Pred. No. 1.5e-10;  
Matches 431; Conservative 175; Mismatches 700; Indels 728; Gaps 97;

```
QY 663 DELLOQHLKMEKERNARKKK-----APAAASEEAAFPVVVEDEMEASGVSGNEE 715
Db 330 EEDVTEKLFDEEDGRSDSEGAEGHDSQSASGEER--PP-----EADGKKGN-- 378
QY 716 EMVEEAEALHAGNEVPRGECGPATVNNSDTE-----SIPSPTEAAKDTGQNGP 767
Db 379 -----SPNSEPTTKTAWAETSRPTEPGPPAPKPLPPPHRGCPAGNMGPGD 427
QY 768 KPAATLGADGPPGPPPTTP-----RRTSRAPTE----- 795
Db 428 YP-----DRGGPPCKPPAPEDEDAWRQRKQSSEISLAVERARRRREERMQEERRA 483
QY 796 -----PTPASEATCAPTPPPAPPSPSPAPPVVPKKEEKEETAAAPPVEEGEE 842
Db 484 ACAEKLRLDEKFGADPKRLKAEPAPPAAPSTAPPAPVKE-----LPAPPA----- 532
QY 843 QRPAAEELAVDTGKAEPEVKSECTEEAEGPAKGDKEAAEAATAEGALKAEEKGGSGR 902
Db 533 --PP-----PASAPTEKEPEEPAQAPQASTPTP-----GVAA 564
QY 903 ATTAKSSGAPQSDSATSADSEVD-----EAEKGDKNLLSPRLLPTTGDPDRANA--- 955
Db 565 APTLVGGGSGTSTSGSFEASVPQLPSKEGPEPEVEVPTTPPVVPKVPKGDGIGP 624
QY 956 -----SPOKPLDLKOLKORAAIPIOTVKVHEPPREDAAATKPAAPPAPPPQ 1003
Db 625 TRQPPSQGLGYPKYOKSLPPRFQROQOOLKQOQOHOHQOQGSAPPTVPVPPSPQPV 684
QY 1004 NLOQESDAPQPGSSPRGSRSPADPAKFAEAAQKLPDPPCWTGCLPPVPVPPREVI 1063
Db 685 TL-----GAVPAKAPPPPKA-----LYPG-----ALGRPPMPMNF- 718
QY 1064 KASHPADPSAFSAPPQHLPLGLUHDTRAPVLPRPTISNPPPLISSAKHPSVLEROIG 1123
Db 719 -----DPRMMNIPYYDP-----RLQGRPPLDIFYPPGV-----HPSGLVPRR 757
QY 1124 ALSQMSVQLHVPYSEHAKAPVPTMGLPMDPKLA-----PESGVKQEO---LSPRG 1176
Db 758 SDSRLSSE---PDRHAPAMLR---ERGTG-PVDP-KLAWGVDVETATPAERPLTSPLR 810
QY 1177 QAGPPESLGVPPTAQEASVLRGTALGSVFGGSITKIGIPSTRVPSDSAITYRGSITHGT 1236
Db 811 QAAEDDKGM-----RSETPPVPP---PPPYLASYPGFPENGTTPG 848
QY 1237 VLYKGTIRI-IGEDSPSLDRGREDSLPKGHVITYEGKGHVLSYEGGMSVTCQSKEDGR 1295
Db 849 -----PISRFLEEGPRL-----PMPGG-----SDEVAK 874
QY 1296 SSSGPPHETAAKRTYDMMEGRVGRAISSAETGLMGRAIIPPERHSPHLLKEQHIRGSI 1355
Db 875 IOTPPKPEKPEETAQLTGPEAGKRLPASR-----SGAGPPPPPRESTETRWGPRGSS 930
QY 1356 TQGIPTSVVEAQEDYLRREAKLLKREGTTP-----PPPPSRDLTBAYTKQALGPLKL 1407
Db 931 RRGIPP-----EPPGAPRRRAGPIKKPPPT-----KVEELPP--- 963
QY 1408 KPAHGLVATVKEAGRSIETPRELRTHTPELAPRLK-----EGSITQ 1453
Db 964 KPLEOG-----DETPK-----PPKPDPLKITKGLGPKETPPNGNLS 1002
QY 1454 GTPKYDTCASTG-----SKKHDRVSLGSPGRTFPVHPDLVDMDADARALE 1500
Db 1003 APRLRDYSYERVGTPSCRGGRGVEFARGFRCTYGRGRG-----AQANSV 1052
```

```
QY 1501 RACYEESLKSPGTASSSGGSIARG--APVIVPELGP-----RQSPLTYVEDHGA 1548
Db 1053 TESFEEM-----GVEVGQGDQITLLPEAALFARHCARVQSMRKSP-----SGA 1097
QY 1549 PFAGHLPRGSPV-----TMREPTRLQEGSLSSKASQDRKLTSTPREIAKSPHSTVP 1601
Db 1098 GSGAQQAARMRVIIWLLQTRRLPHPRREHSPRSS-----RSPTTRSP 1140
QY 1602 EHHPHPISPYEHLLRGVSDVLYRSHIPL-AFDPTSIPIRGIPLDAAAAYYLP-RHLAEN- 1658
Db 1141 TLHR---APARFTCPGV-----ESSLPEGAISPGRREAPPQVPCGMSPPAKSLAPKK 1192
QY 1659 -PTYPHLYPP-----YLIRGPDTAALENRQTIINDYITISQOMHNTATAMAQADML 1710
Db 1193 PPTGP--LPPSKEPLKEKLIPLGPLSPVARGSGNSGVNMGEDGERPRRRHGRHAQQOK- 1249
QY 1711 RGLSPRESSLAL---NYAAGPRGIIIDLSQVPLHPLVLPVPTTGTGPATA---MDRLAYLPTA 1764
Db 1250 ---PPFRRLKQERENAAARGSEG-----KPSLTLPASAPGPEALTJTVA 1292
QY 1765 PQPFSSRHSSPSLPGGPTHLTKPTTSS-----SERERDRDE----- 1803
Db 1293 PAPRAAAKSPDLSNQNSDAQNEEWEETASESDFTSERRGDKEAPPVLLTPKAVGTGPG 1352
QY 1804 -----RDRDREREKSLTSTTTVEHAP IWRPGTEQSGSGG--SSGGGG 1845
Db 1353 GGGGAVPGISAMSGDLSQAKOLSKRSFSQORPQWERNRRPGGKAGSGSGSGGAG 1412
QY 1846 GSSSRPASHSHAHQSPISPTQD--ALQORPSVLHNTGMKGIITAVEPSKPTVLRSTST- 1903
Db 1413 GPGGRTGPGRGDKRSPKNSRSPPEERP-----PGLPLPPPPSSSVFRLDQVI 1464
QY 1904 -SSVVRPAATFPATHCLGTLGCVTYTME-----PVLLPREAPVARPERPRADTG 1956
Db 1465 HSNPA-----GIOQALAQLSRQSVTAPGGHPR-HKPPGPPQAPQG 1504
QY 1957 HAFIAKAPAR-----SGLEPASPSKSGSEPRPLVPVPSGHATITARTPAKNLAPHASP 2009
Db 1505 PS--PRPTRYEPQVNSGL---SSDPHFEPGPMVRGVG-----TPRDSAG---VSP 1550
QY 2010 DPPAPPASADPHREKTKOSKPFSTQLELEL---RSIGYHGSSYSPEGVFVSPVSPSLTH 2066
Db 1551 FPP-----KRERPRPKPELLQESLPPPHSGSGLGS--KPEG---PGQAESRDTG 1597
QY 2067 DKGLPKHL-----EELDKSHLEGELRKPQGPVKLGEEAHLPLRLP-ESOPS 2115
Db 1598 TEALPHIWNRLHTATSRKSYRPSMSEWMEPLSPFEDVAGTEMSQSDSGVDLSGDSQVS 1657
QY 2116 SSPLIQ-TAP--GVKGHORVVTLAQHISEVITQDYTRHHPOQL---SAPLPA-PLYSFPG 2168
Db 1658 SGPCSQRSQSDGGLKG-----AAGGPPKPGGSGSPNLAVPCEGPPG 1698
QY 2169 ASCPVLDLRRPSDLYLPDPDHGAPARGSPHSEGGKRSPEPNKTSVLGGGEDGTEPVSP 2228
Db 1699 SEPP-----RRPP-----PHDGRKELP-----REQPLPPGPIGTERSORT 1736
QY 2229 EGMTEPGHRSRAVYPLLRYDRGDEQTEPSRMGSKSPGNTSOPPAFTSKLTESNAMYKSKQ 2288
Db 1737 DRATEPGPIR-----PSHRPGPPVQF----- 1757
QY 2289 EINKLNTNHNNEPEYNISQGTETFNMPAITGTGLMTYRSQAOVEHASTNGLEAIRK 2348
Db 1758 -----GTSKDSDS-----DLRLVVGDSL-----KAEKE-----LTASVTE 1786
QY 2349 ALMGKYDOWEESPPLSANAFNPLNASLASLPAAMPTIADGRSDHTLTSPGGGGKAKYSGR 2408
Db 1787 AIPVSRD-WELLPSAAASA-EPQSKNLDSDGHCVPPESSSGQRLPYEVFYGSAG----- 1837
QY 2409 PSSRKAKSPAPGLASGRPPSVSVSHSEDCGNRRRTPLTNVRWEDRPPSAGSTPF 2462
Db 1838 PSSSOISGGSHGLS-----ITSKQMLRLRPGTTPSLHPY 1869
```



RESULT 13  
T13049 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13049  
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.  
A:Submitted to the EMBL Data Library, March 1998  
A:Reference number: 217592  
A:Accession: T13049  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2715 <TRE>  
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1  
C:Genetics:  
A:Gene: eld  
A:Cross-references: FlyBase:FBgn0003013  
C:Function:  
A:Description: could act as a transcription factor antagonistic to the Wg pathway  
C:Keywords: DNA binding

Query Match 3.8%; Score 502.5; DB 2; Length 2715;  
Match Local Similarity 20.9%; Pred.No. 2.3e-10;  
Matches 450; Conservative 189; Mismatches 752; Indels 763; Gaps 111;

QY 670 KLMKERNARRKKKAPAA-----ASEEAFPPVVEDEEAEASGVSGNEEMVEEAEA 723  
DB 4 KIKSPQTOOQGGAPAPATPPSAGAPGATPT----- 39

QY 724 LHASNEVPRGCGSPATVNNSSDTEIPS-----PHTAAKDTGONGKPKPATLGA 775  
DB 40 -----SGPPTPNNSNGSDPSIQOQNVAPHYGA-----PPPPGSG-G 77

QY 776 DGPPPGP-----PTPRTSR-----APIEPTASEATGAP----- 806  
DB 78 PGPPPGPAPAAVMHYHLHQOQQHPPPHMQOQHGGAPPPPGGAPEHAPGKEYET 137

QY 807 -TPPPAP-----PSPAPPVVPVVEEKEEATAAPPVEE-----GEE 842  
DB 138 HLPHPHPAYGRYHADPNMDPYRQPLPGKPP-----QQQHPHQOQPPQPGPGSP 193

QY 843 QKPPAAEELAVDTGAEPVKSCTEAEAGP-----AKGDAPAAEATAEGALKAEK 896  
DB 194 NRPPQOQVTPGPPQGGPTTLNLSQSNPPPPQHRYANTYDQOAAASAAAAAQOQ 253

QY 897 EGSGRATTAKSSGAPQSDSATCSADEVDEAEGDKN-----RLLSRP----- 941  
DB 254 QAG-GPPPPGHPGPPPHQOPS-----PYGQOQGWAPPYPRYSPQLGPSQOYRT 301

QY 942 -PSLLTPTGD---PRAN-----ASPQKPLDLKOLKORAAI-----PPIQVTK 980  
DB 302 PPTNTSRGOSYPYPAHQNSGSYSSPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQ 361

QY 981 VHEPPREDAAPTKAPPAP-----PPQNL-----QPESDAPQPGSSPRGK 1022  
DB 362 --OPQOQNTPTTSQYSPYQRYPTPGLPAGSNHRTAYSTHYPEPNRPPGSSPSFG 419

QY 1023 SRSPAPPADKFAEAAQKLPDPCWTSGLFPVPPREVIKASP-HAPDPSAF-SYAPP 1080  
DB 420 SGHPLPLPASP-HVPLPQQOQPPPHVHAGGP---PP-----SSSPHAPSPSPQSP 471

QY 1081 GHPLPL-----GLHD---TARVPLRPPTISNPPPLISSAKHPSVLERIGATS 1126  
DB 472 SPHELIGQNSNDSSGGAHSGMSGPGCTPNPQVMRPTSPPTGSS-----GSR 522

QY 1127 QGMSVOLHVPSYSEAH-----APV---GPVTMGLPLMDPKKLAPFSYKQBLSPRGA 1178  
DB 523 MSPAQAQNHPISTRASNSQSSGGPMQPPVGAGGPPPHPPHGMF-GGPPQOQOQOQQA 581

QY 1179 GPPESLGVPYTAQEAELRGALGTVPGSGITKIGISTRVPDSALITYKSGITHGTADVL 1238  
DB 582 SNSASSANSPOQT-----PPAPPPNQMNMNMTATPPPP---QCAAGGYPMPH 629

QY 1239 YKGTITRIIGEDSPRLDRGREDSLPKGHVIYEGKKHVLVSIEGGMSVTQCSKEDGRSS 1298  
DB 630 MHGGYK--MGPGQSPGAGYPPQPOQ-----QYPPGNYPRPPYPPGAYAT 674

QY 1299 GPHETAAPKRYDWMGVRGVAISSASIEGLMGAIPPE---RHSPHL---KEQHHRG 1353  
DB 675 GPP-----PPPTSQAGAGGANSMPGSAQAGGYPGRGMPNHTGQYPPYQVWPPSPQOQTVPG 729

QY 1354 SITQGIPIRSYVEAQEDYLRREAKLLKREGTTPPPPPSRDLTEAYKTOALGPKLKAHGG 1413  
DB 730 ----GAPGAMVGNH-----VOGKGTPPPPVVG-----GP-----PPPG 760

QY 1414 LVATVKEAGRSIHEIPREELRHTPELAPRLKKGSIQTQGTPLKYDTGASTYGSKKHDV 1473  
DB 761 ----SGSPRLNLY-KQLQHKGGYGGSP-----PQG-PQGYGNG--PTGM----- 800

QY 1474 RSLIGSPGRTFPVH-----PLDVMADALERACYEESLSKSRGCTASSSG 1520  
DB 801 ----HPGMPMPGPHMGPPHGTNNGPPTSTPPQSQMLQ-----GGQPGQCGASGG 847

QY 1521 STARGAPVIVPELGKPROSPLYEDHGAPFAGHLPRGSP-----VTMRP 1565  
DB 848 ----PESGGEH---ISQDNGISSG--PTGAAGMAVTSVVTGPDGTSMDV 892

QY 1566 TPRLOEGSLSSKASQDRKLTSTPREIAKSP-----HSTVPEHHPI-----SPYHL 1614  
DB 893 S---QOSTLSNASASGEDPQCTTPKSRKNDPYQSQHLAPPTSPHVPVHMGPGGGEY 949

QY 1615 LRG-----VSGDLYRSHIPLAFDP---TSIPRGILPDAAYYLPRHLAPNPTYPHYLY 1665  
DB 950 DMSPPNPRPAGSPQVFNHVPVQEPFRSTITTKKSDSLCKLY---EMDDNPD----- 1002

QY 1666 PLYLLRGYD---TALENROTIIINDYITSQOHHNTA---TAAQADMLR-----G 1712  
DB 1003 ----RGWLDKURAFMEERRTPI-----TACTISKQPLDLRLYIYKERG 1045

QY 1713 LSPRESSALNVAAGPRGI-----IDLSQVPHLPVLVPPPT 1748  
DB 1046 FVEVTKSTKWKDIAGLIGASSAAAYTLRKHYNLLTFECHDFRGDIDPLIIOQVEA 1105

QY 1749 GTPATAMDRLAYLTAPOPFSSRHSPLSPGGPHLTKPTTSSSERDRDRDRDR 1808  
DB 1106 GS-----KKTKAASVP-SPGG-GHLDAGTTNSTG----- 1134

QY 1809 EREKSLITSTTVEHAPWIRPGTEOSSGSGSGSGSSSPASHAHQHSIPRTQ 1868  
DB 1135 ----SSNSQDSFPA-PPGSAPNAIDGYPGYGGSPYFGAS-----GPQPDYAT 1178

QY 1869 DALOORPSVLHN--TGKMGIIITAVEPSKPTVLRSTSSSPVR-PAATFPATHCPGLGTL 1925  
DB 1179 AQOMQRPSONNPQTPHGAANAAGD-----NISVNPFDPTAAG-----GGPG 1225

QY 1926 DGYYPTLMEVLLPKEAPVARPERPRADTHAFIAKPPARSGLEPASSPSKSGEPRLV 1985  
DB 1226 SGTGPG-----PQGP-----GPGAASGA-----GAVGAVGGGPQHP 1259

QY 1986 PVSCHAITARTPA---KNLAPHASPPPPAPPASADPHREKTQSKPSIQELESRLG 2042  
DB 1260 PPHPSHTAAQAQOHOQOHPQHGLPGPPPPQOQOQOQOQOQOQOQOQOQOQOQ 1310

QY 2043 YHGSYSPE---GYEVPSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKOPGVKLG 2098  
DB 1311 -CGPPPAQOQHGQGVPPSP-----QOHVPAAGAYPPPG 1345

QY 2099 EAAHLPLRLPESQSSPPLLQTAGVKGHQVVTVAQHISEVITQDTRHHPOOLSAP 2158  
DB 1346 SG-----YPTVSVTSPGSP-YPSQPGAYG-----QYSSDOYNATGP 1381

QY 2159 LPAPLYSPFGACPVLDLRPPSDLYL-PP--PDHGAAPARGS-----PHSEGGKRS 2206  
DB 1382 PGQPPGQPG-----QYPPQNRMVPPYGEAPGTGANQYGPYGRSPYSPPPG 1433











Db 695 ---EALGRLDGLFGQKAPKETAFTSLSSLATPAGATEHVSYIQDETIPGYSETEQTISDE 751  
QY 526 EKEAEKEEB--KPEVENDKEDLLKKTDDTSGEDNDEKAVASKRKTANSQGR--K 579  
Db 752 EIHDEDEPARPREPTSTYDLSGPEGCPFFASQADSAVFASSKTYGAPETELTYPP 811  
QY 580 GRITRSMANEANSEAITPOQSAELASM-----ELNESSRTEEMETAKKLLHEHGRN 633  
Db 812 NMVAAPLAEHEHVSSATSITCDKLSSFATVAEDQSVAASLTAPQTEETGKSSLL----- 866  
QY 634 WSAIARMGSKTVSOCKNFFYKRONLDEILOQHKLAMEKERNARRKKKAPAAASEE 693  
Db 867 LDTVTSIPSRT-----EATQGLDYVPSAGTI-----SPTSSLEE 901  
QY 694 ---AAPPVVEDEMEBASVGSNEEREMVEEAALHASGNVPRGECSPATVNNSSDTE 750  
Db 902 DKGFKSP-----CEDFSVTESEK-----KGETVGRG--LSGKAVGKEEYV- 943  
QY 751 IPSPHTEAAKDTGONGPKPATLGADGP--PPGPTTPPRTSRAPTEPTPASEATGAPT 808  
Db 944 ----VTSEKLSQYA----AVFGAPGHTLPPGEPALGEVEERC--LSPDDSTVKMASP--P 992  
QY 809 PPAPSPSAPP-----PVVPKEKEE-----ETAAAPPV--EAGEEQK--PAAABELAVD 854  
Db 993 PSGPSSAAHTPHQSVEDKSEPRFQEDSWGETHSPGVSKEDSEEQTVKPGPEEGTSE 1052  
QY 855 TGKAEPVVKSECTEAEAGPAKGADEAAEATAE--GALKAEKKEGGG-----R 902  
Db 1053 EKGK--PPTSPQAOAMPVSIAGGOTCTIQLLPEQDKAIVFETGAGSNLGAITLPGEV 1111  
QY 903 ATAKSSGAPQSDSATSADBEVDAEGDKNRLSPRSLTTPGDPANASQPKPLD 962  
Db 1112 TSTEEATEPKQDEVLRTDQSLSPEDAESLSVSVSPDT-----KQEATPRSPCS 1163  
QY 963 LKOLKORAAAIPIQVTKVHEPREDAATKAPAPPPQNLQEPDAPQPGSSPRCK 1022  
Db 1164 LKEQHQKDLNMPVEDPQSLSFSESPSKET--SLDISKQLSPSLGTQFQELNLGK 1222  
QY 1023 -SRSPAPPADKEAFAEAQKLPDPCWTSGLFPVPPREVAKAS-----PHAPDP--SAFS 1076  
Db 1223 EERGPVMKAE-----DDSCHLAPVSIPEPHRATVSPSDETPAGTLPGGSFS 1269  
QY 1077 YA-----PGHPLPLGLH--DTARPVLPRPTISNP-----PP 1107  
Db 1270 HSALSVDRKHSPGEITPGGHEFWTSDSLTKSPESLSSPAMEDLAVEWEGKAPKEKEPE 1329  
QY 1108 LISSAKHP--SVLERQIGAISQMSV-----QLHVPISEHAKP----- 1144  
Db 1330 LKSETRQKQKQILPEKVAVVEQDILIHQKDGALDENKPGROOKTPEQKGRDLDEKDTA 1389  
QY 1145 ----VGPVTMGLPLMDPKKLAPFGVKQEQLSPRGQ--AGPPESLGVPTAQEASVLRGTA 1199  
Db 1390 AELDKP-----EPKE-----KDLREDQQRAGPAE-----KDKASQORDTD 1428  
QY 1200 LGSVPGGSITKIPSTRVPSDSAITYRG-----ITHGTPADVLYKGTITRIIGED- 1250  
Db 1429 LQOT-----QATEPRDAQERRDSEKDKSLRLDRTPPE--KDRI--LVQEDR 1473  
QY 1251 ----SPSRLDR-----GREDSLPGHVIYE-----GKKGHVLSYEG--GM 1284  
Db 1474 APEHSIPEPTQDRADPRKGTDDKQKEEASEEQVLEQDKWALGKEGETLDQEARTAE 1533  
QY 1285 SYTQSKED---GRSSSGPPHETAAPKRTY-----DMMEGRVGRAISSASTEGLMG 1332  
Db 1534 QKDETLKEDTKOGKSSFFVEDTTSKETSIVLDQKSAEKADSVEQDQGALEKTRALGL-- 1591  
QY 1333 RAIPPE-----RHSPLHLEKQHIIRG-----SITQIGIPRSYVEAQEDYLRRKAKLLKREG 1383  
Db 1592 EESPAEGSKAREQEKYKWEQDVQGWRETSPTRGEP--VGGQKE-----PVPAWEK 1642  
QY 1384 PPPPP--PSRLDT--EAYTQALGPLKLKPAHE-----GLVAIVKAGRS--THEIPR 1430  
Db 1643 SPEQEVRYWRDRDITLQODAYVREILSCDRKWNFPHELDQGARPRYCEBERESTFLDEGPD 1702

QY 1431 EE---LRHTPELPLAP--RPLKEGSIQTGTPLKYDGTGASTTGSKKKHDVRSLLIGSPGRTF 1484  
Db 1703 EGEITPLQHTPRSPWTSDFKDFQEPQLPKQGLEVERWLAEVGLPPEEDKLTTRSPFEI 1762  
QY 1485 -PPVHPLDVMADARALACRYEES-----LKSRRPGTAS-----S 1517  
Db 1763 SPPASPPEN--TGQVRPSAPGQESVPDPTSTAPMRNEPTTSPWLAETPPWVPKDRPLPP 1820  
QY 1518 SGGSTARGAPVIVPELGPQSPPLY-----EDHGAPFAGHLPRGSPV 1560  
Db 1821 APLSPAPAPPTPAPPHPT--VPFSGWLAEDYSVVAQVEGAEELEGGYS--PLGKDY 1875  
QY 1561 TWREPTPRLOEGSSSKASODRKLTTSTPRETA--KSPHSTVPEHHPHPTISPYEHLRLGVS 1619  
Db 1876 RKAEGERGEAGAPDSSSPKVPKEAGESLATRDTEQTEPEQREPTYPDE---RSFQ 1932  
QY 1620 GVDLYRSHIPLAFDPTSIPIGIPDAAAAYILPRHLA-----PNPTYPH 1663  
Db 1933 YADIYEQMMLTGLGPACTPREPLGASGDW--PPLHSTKEEAAGCNTSAEKETSSPASQ 1990  
QY 1664 -----LYPPYLIRGYPDYA-----ALENQTIINDYITISQOMH 1696  
Db 1991 NLOSTPAFASYASLAGPAVPP--ROEPDGPVNVBSITPPAVPPRAPISLKDLSPLN 2047  
QY 1697 HNTATAMAQR-----ADMRLGL-----SPRE----- 1717  
Db 2048 GSTVSCSPDRRTTSPKETCGRHWDGTDNDSDLEKAREQPEKETRSPSPHHPMWMGHSSL 2107  
QY 1718 -----SSL-----ALNYAAGRGIIDISQVPHLPVLVPPPTGTATAM---- 1755  
Db 2108 WPETEAYSSLDSSHLGVSVRPSLDPPASAFGSLLQAP--POL--PSPAEPRSPACGSL 2163  
QY 1756 ---DR-LAYLTPAQPTSSRHSS-----SPLSPGCP--THLTKPTT 1790  
Db 2164 AFSGRDALALVPGTPT--RTRHDEKLVTKAPSLDSSLPOLPSPSPGGLLNLPKPRAS 2221  
QY 1791 TSSSERDRDRDRDREREKSIITTTTVEHAPIWRPTEQSSGSSGSSGGSSSR 1850  
Db 2222 PALSE-----GSSEATTPIVSSVAERFP--PGLE---AAEQSAEGLSGKE 2263  
QY 1851 PASHSHAHQHSIPISPTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVRPA 1910  
Db 2264 SAHHS--LMDLTPLSPASASLDLAPA-----PA 2290  
QY 1911 ATFFPATHCPLGTLGDVYPTLME-----PVLLPK-----EAPR-----V 1945  
Db 2291 PAPAPAPGLP--GDLGDGTLPCRPECTGELTKKPSPLSPSGDHEANGPGETSLNPPGFVT 2349  
QY 1946 ARPERPRADTGHAF-----LAKPPARSGLPEPASSPKSGSEPRPLVPPVSGHATI 1994  
Db 2350 ATAEEAEAEAPHAWERGSWPEGAERSSRPDTLLSSEQPLRPGKSSGG---PPCSLSSEV 2405  
QY 1995 ARTP--AKNLAPH--HASP---DPPAPPASADPHREKTSQKPFISIOLEL-----RS 2040  
Db 2406 EAGQGCATDPRHCGELSPSLNPLPPS-----TODSDLSTEARELAGKGRRR 2456  
QY 2041 LGYHCSSYSP---EGVEPVSPVSPSLTHDKGLPKHLEELDLSHLEGLRLRQPG---P 2093  
Db 2457 VGRGATGCPMADETPTTSAOSSSDSDVPPTEECPSITAEEALDSEDEGDFLP 2516  
QY 2094 V-KLGG-----EAHLPHRLPSPSSPQLLTQAPGVKGQRVVTLAQIHSEVITQ 2145  
Db 2517 VDKAGVSGTHHPRGHDPPTPLDPRSP--RP 2550  
QY 2146 DYTRHHPOOLSAPLAPLYSPFGASCPVLDLRRPSLDYLPDPDHGAPARGSPHSEG--- 2202  
Db 2551 DVCMDAPBGLSS-----ESGRVERLREKR-----PGRRAPGRAPKAPSPARRL 2593  
QY 2203 ---GRSPPEPNKTSVLGGEDGIEPVSPPEGMT---EPGH---SRSAVYPLLIRDGQOT 2252  
Db 2594 DIRGKRSTPGKGVDRRTSRTVPRPRSTPSQVTSAEKDGHSFMSKGLV-----NLGA 2647











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 8, 2001, 14:32:18 ; Search time 53.39 seconds  
(without alignments)  
1614.929 Million cell updates/sec

Title: US-09-522-753-5  
Perfect score: 13215  
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEEPRKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	13215	100.0	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
2	10987.5	83.1	2472	1 NCR2_MOUSE	Q9W442 mus musculus
3	4187	31.7	2453	1 NCR1_MOUSE	Q60974 mus musculus
4	4147.5	31.4	2440	1 NCR1_HUMAN	O75376 homo sapien
5	714.5	5.4	533	1 NCR1_RAT	Q9WUB5 rattus norv
6	526	4.0	2142	1 BAT2_HUMAN	P48634 homo sapien
7	492.5	3.7	2774	1 MAPA_RAT	P34926 rattus norv
8	443.5	3.4	2805	1 MAPA_HUMAN	P78559 homo sapien
9	428.5	3.2	2715	1 TRX2_HUMAN	Q9UMN6 homo sapien
10	424.5	3.2	2464	1 MAPB_MOUSE	P14873 mus musculus
11	412.5	3.1	1226	1 YCS3_YEAST	P25357 saccharomyc
12	410	3.1	2468	1 MAPB_HUMAN	P46821 homo sapien
13	400.5	3.0	1464	1 CAL1_HUMAN	P02452 homo sapien
14	391.5	3.0	1411	1 TCOF_HUMAN	Q13428 homo sapien
15	391.5	3.0	1460	1 CAL1_CANFA	Q9XSJ7 canis faml
16	391	3.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
17	390	3.0	2944	1 CAL7_HUMAN	Q02388 homo sapien
18	389.5	2.9	3866	1 HRX_MOUSE	P55200 mus musculus
19	389	2.9	1670	1 CA34_HUMAN	Q01955 homo sapien
20	380	2.9	3669	1 HRX_HUMAN	Q03164 homo sapien
21	379	2.9	1685	1 CA54_HUMAN	P29400 homo sapien
22	378	2.9	1183	1 DRPL_RAT	P54258 rattus norv
23	373	2.8	1763	1 CA24_ASCSU	P27393 ascaris suu
24	367	2.8	1185	1 DRPL_HUMAN	P54259 homo sapien
25	358	2.7	1618	1 NEST_HUMAN	P46881 homo sapien
26	355.5	2.7	2842	1 APC_RAT	P70478 rattus norv
27	355	2.7	2090	1 N214_HUMAN	P35658 homo sapien
28	353	2.7	1453	1 CAL1_CHICK	P02457 gallus gall
29	352	2.7	1838	1 CAL5_HUMAN	P20908 homo sapien
30	350.5	2.7	2843	1 APC_HUMAN	P25054 homo sapien
31	348.5	2.6	1464	1 CAL3_MOUSE	P08121 mus musculus
32	348	2.6	1669	1 CAL14_MOUSE	P02463 mus musculus
33	348	2.6	3256	1 K167_HUMAN	P46013 homo sapien

## RESULT 1

ID	NCR2_HUMAN	STANDARD;	PRT;	2517 AA.
AC	Q9Y618; Q9Y5U0; O13354; O00613; O15416;			
DT	01-OCT-2000 (Rel. 40, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF			
DE	RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (SMRTE) (THYROID-			
DE	RETINOIC ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-			
DE	ASSOCIATING FACTOR) (TRAC) (CTG26).			
GN	NCOR2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM SMRT).			
RC	TISSUE=Pituitary;			
RX	MEDLINE=99178941; PubMed=10077563;			
RA	Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;			
RT	"Unique forms of human and mouse nuclear receptor corepressor SMRT.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM SMRT).			
RC	TISSUE=Cervix adenocarcinoma;			
RX	MEDLINE=99199215; PubMed=10097068;			
RA	Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;			
RT	"SMRTE, a silencing mediator for retinoid and thyroid hormone			
RT	receptors-extended isoform that is more related to the nuclear			
RT	receptor corepressor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).			
RN	[3]			
RP	SEQUENCE OF 1023-2517 FROM N.A.			
RC	TISSUE=Cervix adenocarcinoma;			
RX	MEDLINE=96008552; PubMed=7566127;			
RA	Chen J.D., Evans R.M.;			
RT	"A transcriptional co-repressor that interacts with nuclear hormone			
RL	Nature 377:454-457(1995).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM TRAC-1).			
RC	TISSUE=Fetal liver;			
RX	MEDLINE=96408715; PubMed=8813722;			
RA	Sande S., Privalsky M.L.;			
RT	"Identification of TRACs (T3 receptor-associating cofactors), a family			
RT	of cofactors that associate with, and modulate the activity of,			
RT	nuclear hormone receptors.";			
RL	Mol. Endocrinol. 10:813-825(1996).			
RN	[5]			
RP	SEQUENCE OF 428-613 FROM N.A.			
RC	TISSUE=Brain cortex;			
RX	MEDLINE=97369492; PubMed=9225980;			
RA	Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,			
RA	Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;			
RT	"CDNAs with long CAG trinucleotide repeats from human brain.";			

P34333 caenorhabdi  
Q61315 mus musculu  
P12036 homo sapien  
P11087 mus musculu  
P02461 homo sapien  
P53420 homo sapien  
P13942 homo sapien  
P39060 homo sapien  
O14497 homo sapien  
P02458 homo sapien  
P12270 homo sapien  
P08640 saccharomyc

## ALIGNMENTS







QY 841 EEOKPPAAEELAVDTGKABEPVYKSECTEABEGPAKGDAAEATAEAGAKAEKKEGS 900  
DB 841 EEOKPPAAEELAVDTGKABEPVYKSECTEABEGPAKGDAAEATAEAGAKAEKKEGS 900  
QY 901 GRATTAKSSGAPQDSSATCSADEVEAEGDKNRLSPRSLTPTGDPNANASPOKP 960  
DB 901 GRATTAKSSGAPQDSSATCSADEVEAEGDKNRLSPRSLTPTGDPNANASPOKP 960  
QY 961 LDLKOLKORAAAIPTQVTKVHEPPREDAAPTKPAAPPQPPQNLQPEADAPQPGSSPR 1020  
DB 961 LDLKOLKORAAAIPTQVTKVHEPPREDAAPTKPAAPPQPPQNLQPEADAPQPGSSPR 1020  
QY 1021 GKSRSAPPADKEAFAAEAAQKLPDGPCCWTSGLPVPVPREVIKASPHADPSAFSAPP 1080  
DB 1021 GKSRSAPPADKEAFAAEAAQKLPDGPCCWTSGLPVPVPREVIKASPHADPSAFSAPP 1080  
QY 1081 GHPLPLGLHDTARVPLPREPTISNPPPLISSAKHPSVLERQIGASQGMVQLHVPYSEH 1140  
DB 1081 GHPLPLGLHDTARVPLPREPTISNPPPLISSAKHPSVLERQIGASQGMVQLHVPYSEH 1140  
QY 1141 AKAPVGVMTGLPLPMDPKKLAFFSGVKQEQLSPRGQAGPPESGLVPTAQEAASVLRGTAL 1200  
DB 1141 AKAPVGVMTGLPLPMDPKKLAFFSGVKQEQLSPRGQAGPPESGLVPTAQEAASVLRGTAL 1200  
QY 1201 GSVPGGSIKGIPTSTRVPSDSAITYKGSITHTGTPADVLYKGTITRIIGEDSPSLDRGRE 1260  
DB 1201 GSVPGGSIKGIPTSTRVPSDSAITYKGSITHTGTPADVLYKGTITRIIGEDSPSLDRGRE 1260  
QY 1261 DSIKPKGHVLYECKKGVLSYEGGMSVTQCSKEDGRSSGPPHETAPKRTYDMMEGRVGR 1320  
DB 1261 DSIKPKGHVLYECKKGVLSYEGGMSVTQCSKEDGRSSGPPHETAPKRTYDMMEGRVGR 1320  
QY 1321 AISSASIEGLMGRAIPPERHSHLKEQHIRGSIITQIPRSYVEAQEDYLREAKLLK 1380  
DB 1321 AISSASIEGLMGRAIPPERHSHLKEQHIRGSIITQIPRSYVEAQEDYLREAKLLK 1380  
QY 1381 EGTPTPPPPSRDLTAYKTOALGPKLPAHEGLVATVKEAGRSIHEIPREELRHTPELP 1440  
DB 1381 EGTPTPPPPSRDLTAYKTOALGPKLPAHEGLVATVKEAGRSIHEIPREELRHTPELP 1440  
QY 1441 LAPRLKEGSIITQGTPLKYDTGASTGSKKHVDVRSIIGSPGTFPPVHPDVMADARALE 1500  
DB 1441 LAPRLKEGSIITQGTPLKYDTGASTGSKKHVDVRSIIGSPGTFPPVHPDVMADARALE 1500  
QY 1501 RACYEESLSRPGCTASSSGSITARGAPVIVPELGPQRQSLTYEDHGAFAGHLPRGSPV 1560  
DB 1501 RACYEESLSRPGCTASSSGSITARGAPVIVPELGPQRQSLTYEDHGAFAGHLPRGSPV 1560  
QY 1561 TWREPTPRLOEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHIPISPYEHLRGVSG 1620  
DB 1561 TWREPTPRLOEGSLSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHIPISPYEHLRGVSG 1620  
QY 1621 VDLYRSHIPLAFDPTSIPIRGIPDLDAAYYLRHLAPNPTYPHLYPPYLRGYPDTAALE 1680  
DB 1621 VDLYRSHIPLAFDPTSIPIRGIPDLDAAYYLRHLAPNPTYPHLYPPYLRGYPDTAALE 1680  
QY 1681 NQOTIINDYITSOQMHNTATAMORADMRLGLSPRESSLALNYAAGPRGIIDLSQVPHL 1740  
DB 1681 NQOTIINDYITSOQMHNTATAMORADMRLGLSPRESSLALNYAAGPRGIIDLSQVPHL 1740  
QY 1741 PVLVPTPTCTATAMDRLAYLPTAQPFFSRHSSSPLSPGGPHTLTKTPTTSSSERERDR 1800  
DB 1741 PVLVPTPTCTATAMDRLAYLPTAQPFFSRHSSSPLSPGGPHTLTKTPTTSSSERERDR 1800  
QY 1801 DRERDREREREKSIITSTTVSHAPVWRPGTEQSSGSGSGSGSGSSSRPASHAHQH 1860  
DB 1801 DRERDREREREKSIITSTTVSHAPVWRPGTEQSSGSGSGSGSGSSSRPASHAHQH 1860  
QY 1861 SPISPRTDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSSPVREAAFPPTATHCP 1920  
DB 1861 SPISPRTDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSSPVREAAFPPTATHCP 1920  
QY 1921 LGGTLGVYPTLMEPVLLKPEARVPRADTGHAFLLAKPPARSGLPASPSSKGSE 1980

DB 1921 LGGTLGVYPTLMEPVLLKPEARVPRADTGHAFLLAKPPARSGLPASPSSKGSE 1980  
QY 1981 PRPLPVVPSGHATIAITPAKNLAPHASPDPPAPPASADPHREKTOSKPFSTQLELRS 2040  
DB 1981 PRPLPVVPSGHATIAITPAKNLAPHASPDPPAPPASADPHREKTOSKPFSTQLELRS 2040  
QY 2041 LGYHGSSYSPGEGVEPVSPVSSSLTHDKGLPKHLELDKSHLEGELRPKQPGVKLGGEA 2100  
DB 2041 LGYHGSSYSPGEGVEPVSPVSSSLTHDKGLPKHLELDKSHLEGELRPKQPGVKLGGEA 2100  
QY 2101 AHLPHLRPLPESQSPSSPLLQTPAGVKGHQVVYTLAQHISEVITQDTRHHPOOLSAPLP 2160  
DB 2101 AHLPHLRPLPESQSPSSPLLQTPAGVKGHQVVYTLAQHISEVITQDTRHHPOOLSAPLP 2160  
QY 2161 APLYSPFGASCVPDLRRPPSDLYLPPDPHGPARGSPHSEGGKRSPEPNKTSVLGGED 2220  
DB 2161 APLYSPFGASCVPDLRRPPSDLYLPPDPHGPARGSPHSEGGKRSPEPNKTSVLGGED 2220  
QY 2221 GIEPVSPPEGMEPTGHSRSVAVPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESNS 2280  
DB 2221 GIEPVSPPEGMEPTGHSRSVAVPLLYRDGEQTEPSRMGSKSPGNTSQPPAFFSKLTESNS 2280  
QY 2281 AMVSKKKQEKINLKNTHNNEPEYNISQPGTEIFNMPAITGTGLMTYRSQAVOEHAHN 2340  
DB 2281 AMVSKKKQEKINLKNTHNNEPEYNISQPGTEIFNMPAITGTGLMTYRSQAVOEHAHN 2340  
QY 2341 GLEAITIRKALMGKYDQWEESSPPLSANAFNPLNASASIPAAAMPITAADGRSDHDTLTPGGG 2400  
DB 2341 GLEAITIRKALMGKYDQWEESSPPLSANAFNPLNASASIPAAAMPITAADGRSDHDTLTPGGG 2400  
QY 2401 GRKAVSGRPSRRKAKSPAPGLASGDRPPSVSVHSGDCNRRRTPLTNRVWEDRPPSSAGST 2460  
DB 2401 GRKAVSGRPSRRKAKSPAPGLASGDRPPSVSVHSGDCNRRRTPLTNRVWEDRPPSSAGST 2460  
QY 2461 PFPYNPLINRLAQVYMASPPPPGLPAGSGPLAGPHAWDEEPKPLCSQYETLSDSE 2517  
DB 2461 PFPYNPLINRLAQVYMASPPPPGLPAGSGPLAGPHAWDEEPKPLCSQYETLSDSE 2517  
RESULT 2  
NCR2\_MOUSE STANDARD; PRT: 2472 AA.  
AC Q9WU42; Q9WU43; Q9WU41;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF  
DE RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (SMRTE) (THYROID-,  
DE RETINOIC-ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-  
DE ASSOCIATING FACTOR) (TRAC).  
GN NCOR2 OR SMRT  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A. (ISOFORM ALPHA AND ISOFORM BETA).  
RC TISSUE=Spleen, and Brain;  
RX MEDLINE=99178941; PubMed=10077563;  
RA Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;  
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
RN [2]  
SEQUENCE FROM N.A. (ISOFORM ALPHA).  
RC TISSUE=Embryo;  
RX MEDLINE=99199215; PubMed=10097068;  
RA Park E.J., Schreien D.J., Yang M., Li H., Li L., Chen J.D.;  
RT "SMRTE, a silencing mediator for retinoid and thyroid hormone  
RT receptors-extended isoform that is more related to the nuclear  
RT receptor corepressor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
CC !- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME







Db 717 ARASQAGNEVPRVGCSPGAANNSSDYESPRSEATKDT---GPKPTGTEALPAAT 773  
QY 778 PPPGPTTTPRRTSRAPTEPTAPSEATGAPTPPPAPSPSAPP-----VVP 823  
Db 774 QPVP--PPEEFAVAPAESPVDPASGPPSP--EPSSHILPHRLLWTRMKNKPRLLQLP 828  
QY 824 KEEKEETANAAPVEGEBQKPAEELAVDTGKAEPVKSECTEAEAGPA-KGKDAEA 882  
Db 829 RORMPRSRLRPRRSMWKEPEPEASE-----EPPEVKSDHKEETEPEDEKAGTEA 882  
QY 883 AEATAEGALKAEEKEGGCRATT-AKSSCAPODSSSATSADADEABEGGDKNLLSPR 941  
Db 883 IETVSEAPLKV-EAGSKAAATKSSGATQDSSSATSADADEPEGGDKGLLSPR 940  
QY 942 PSLLPTGDPANASQPKPLDLKQLKORAAAIPIQVTKVHEPPPREDAAPTAPAPPAPP 1001  
Db 941 PSLLPTAGDPRASTSPQKPLDLKQLKORAAAIPII-VTKVHEPPREDIVPKPVPVPPPP 999  
QY 1002 PONLOPESDAPQOQSSSPRGKRSRAPPADKE-----AFAEAQKLPDPPPCWTSGL 1053  
Db 1000 TQHLQPEGDVSQOQSSSPRGKRSRVPVPAEKAEPAPFPAPTEGPKLPTPEPRWSSGL 1059  
QY 1054 PFPVPPREVIKASPHADPSAFSPAPGHPPLGLGHDARTARVLPRTTNSNPPPLISSAK 1113  
Db 1060 PFPPIPPREVIKTSPPHAADPSAFSPYPPGHPPLGLGHDARSAPVLPAPP-ISPNNPPLISSAK 1118  
QY 1114 HPSVLEROIGATS-OGMSVOLHVPVSEIAKAPVPMGLPLPMDPKKLAIPSGVKQEOL 1172  
Db 1119 HPGVLERLQALISQOQMSVOLRVPHSEHAKAPMGPLTGMGLPLAVDPKKL----- 1167  
QY 1173 SPRGOAGPESLGVPTAQEASYLRTALGSVPGSITTKGIPSTRVPSDAITYRGSITHG 1232  
Db 1168 -----GTALGSATSGSITKGLPSTRAADGP-SYRGSITHG 1201  
QY 1233 TPADVLYKGTITRIIGEDSPSLDRGREDSPKGVHIVTEGKKGHVLSYEGGMSVTCSCKE 1292  
Db 1202 TPADVLYKGTISRIYGEDSPSLDRAREDTPKGVHIVTEGKKGHVLSYEGGMSVTCSCKE 1261  
QY 1293 DGRSSGPPHETAAKRTYDMMEGRVGRRAISSATEGLMGRAIPPERHSPHLKQOHILR 1352  
Db 1262 DGRSSGPPHETAAKRTYDMMEGRVGRVTVTSASIEGLMGRAI-PEQHS-PLKQOHILR 1319  
QY 1353 GSITQIGIPRSYEAQEDYLRLREAKLLKREGTPPPPPPSRDLTEAYKTQ---ALGPLKLP 1409  
Db 1320 GSITQIGIPRSYEAQEDYLRLREAKLLKREGTPPPPPPPRDLTEYTKPRPLDPLGLPLKLP 1379  
QY 1410 ANEGLVATVKEAGRSIHETIPRELHRTPELPLAPRLKEGSTTQGTPLKYDTGASTTSGK 1469  
Db 1380 THEGVVATVKEAGRSIHETIPRELHRTPELPLAPRLKEGSTTQGTPLKYDSCAPSTGK 1439  
QY 1470 KHDVRSLSIGSGRTFPPVPHPLDMDADARALACRYEESLKSRRGTASSSGGSIARGAPVI 1529  
Db 1440 KHDVRSIGSGRTPFALHPLDMDADARALACRYEESLKSRRGTASSSGGSIARGAPVI 1499  
QY 1530 VPGLKQPSPLTYEDHGAPFAGHLPRGSPVMTREPTPRLOEGSLSSKASQDRKLSTP 1589  
Db 1500 VPGLKQPSPLTYEDHGAPFAGHLPRGSPVMTREPTPRLOEGSLSSKASQDRKLSTP 1559  
QY 1590 REIAKSPHSTVPEHHPHPLISPEHLLRGVSGVDLYRSHIPLAFDPTSPRGIPLD-AAAA 1648  
Db 1560 REIAKSPHSTVPEHHPHPLISPEHLLRGVSGVDLYRSHIPLAFDPTSPRGIPLD-AAAA 1619  
QY 1649 YLPLRHLAPNPTYPHLYPYLLIRGYPDTAALENROTINDYITSOOMHNTATANAQAD 1708  
Db 1620 YLPLRHLAPNPTYPHLYPYLLIRGYPDTAALENROTINDYITSOOMHNTATANAQAD 1679  
QY 1709 MLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPPPTGPTATAMDRLAYLPTAPOPF 1768  
Db 1680 MLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPPPTGPTATAMDRLAYLPTAPOPF 1739  
QY 1769 SSRHSSSSPLSGGPHLTKPTTSSSERDRDRDRDRDREREKRSILSTTTTVEHAPIWR 1828  
|||||

Db 1740 SSRHSSSSPLSGGPHLAKPTATSSSERERERERD-----KSILTSTTTVEHAPIWR 1793  
QY 1829 PTEQSSSSSGSGGGGSSRSPASHAHQSPISPRQDALQQRPSVLHNTGMGIIT 1888  
Db 1794 PTEQSSGA-----GGSSRPASHT--HQSPISPRQDALQQRPSVLHNTSMGVVT 1843  
QY 1889 AVEPKSPTVL-----RSTSTSSPVRAATPPATHCPILGGTLDGVYPTLMEPVLLPKBAPRV 1945  
Db 1844 SYEPGPTVLRWARSTSTSSPVRAATPPATHCPILGGTLDGVYPTLMEPVLLPKBAPRV 1903  
QY 1946 ARPERPRADTGHAFIAPKPARSCLEPASPSPKSGSEPRPLVPPVSGHATARTAPAKNLAPH 2005  
Db 1904 ARPERARVDAGHAFITKPPGR---EPASSPSKSEPRSLAPSSSSHTARTAPAKNLAPH 1960  
QY 2006 HASPPPPAPPASADPHREKTSQKPFISQIELELSLGYH-GSSYSPGEGVEPVSPVSSPSL 2064  
Db 1961 HASPDPPA-PTSASDLHREKTSQKPFISQIELELSLGYHSGAGYSPDVEPISPVSSPSL 2019  
QY 2065 THDKGLPKHLEBDKSHLEGELRPQKPGVKILGGEAAHILPHLRPLPDESQFSSPLLQATP 2124  
Db 2020 THDKGLSKPLEELEKSHLEGELRHKQKPGPMKLSAEAAHILPHLRPLPDESQFSSPLLQATP 2079  
QY 2125 GYKGHORVVTLAQHISEVITQDYTRHHPOOLSAPLPAPLYSPGASCPVLDLRRPPSDLY 2184  
Db 2080 GIKGHORVVTLAQHISEVITQDYTRHHPOOLSGPLPAPLYSPGASCPVLDLRRPPSDLY 2139  
QY 2185 LPPPHGAPARGSPHSEGGKRSPEPKNTSVLGGGEGDIEPVSPPEGMEPEGHSRSAYVPL 2244  
Db 2140 LPPPHGTPARGSPHSEGGKRSPEPKNTSVLGGGEGDIEPVSPPEGMEPEGHARSTAYPL 2199  
QY 2245 LYRDGEQTEPSRMGSKSGPNTSQPAFFSKLTESNAMYKSKQKIEINKLNTNHNREPEY 2304  
Db 2200 LYRDGEQTEPSRMGSKSGPNTSQPAFFSKLTESNAMYKSKQKIEINKLNTNHNREPEY 2258  
QY 2305 NISQGTETEFNPAITGTMYSRQAOVQEHASTNMGLEAIIKALMGKYDQWEEPPPLS 2364  
Db 2259 NISQGTETEFNPAITGTMYSRQAOVQEHASTNMGLEAIIKALMGKYDQWEEPPPLS 2318  
QY 2365 ANAFNPLNASLIP-AAMPITADGRSDHILTSPPGGGKAKYSGRPSRKKAKSPAPGLAS 2423  
Db 2319 ANAFNPLNASLIP-AAMPITADGRSDHILTSPPGGGKAKYSGRPSRKKAKSPAPGLAS 2378  
QY 2424 GDRPPSVSVHSEGDGNCNRRTPLTNRWEDRPSAGSTPEPPYPLNRLQAGVNASPPPG 2483  
Db 2379 GDRPPSVSVHSEGDGNCNRRTPLTNRWEDRPSAGSTPEPPYPLNRLQAGVNASPPPG 2438  
QY 2484 LPAGSGPLAGPHANDEEPKPLLCQYETLSDSE 2517  
Db 2439 LAAGSGPLAGPHANDEEPKPLLCQYETLSDSE 2472  
RESULT 3  
NCRL\_MOUSE  
ID NCRL\_MOUSE STANDARD; PRT; 2453 AA.  
AC Q60974; Q60812;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (RETINOID X RECEPTOR  
DE INTERACTING PROTEIN 13) (RIP13).  
GN NCOR1 OR RXRIP13.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCRL\_Faxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
RC TISSUE=Pituitary;  
RX MEDLINE=96008539; PubMed=7566114;  
RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B.,  
RA Kurokawa R., Ryan A., Kamei Y., Soederstroem M., Glass C.K.,  
RA Rosenfeld M.G.;  
RT "Ligand-independent repression by the thyroid hormone receptor







```
Db 961 PIGTPVSGYALYORIKAMHESALLE-----EQROEQVDLECRSSTSPCST 1008
QY 1023 SRSPAPPADKEAFAAEAAQKLPDPCWTSGLPFPVPPPREVIKASHPADPSAFSAPPCH 1082
Db 1009 SKSP-----NRE-----W-----EVLQAPH-----QVITNLPEGV 1034
QY 1083 PLPLGLHDTARVLPRLPPTISNPPPLISSAKHPSVLEQIGAIISQ-----MSVOLHVPY 1137
Db 1035 RLP-----TTRTPRPPLIPSSKTTVASEK-PSFI--MGGSISQGTGCTYLSSHQAAPP 1086
QY 1138 SEHAKAPGVPTMGLPLMDPKLAPFSCVKQEOQLSPRGCAQPPESLGVPTAQEAASVLRG 1197
Db 1087 QAPKPSVGSISLGLPQOESTKAAPLTIKOEFSRPSQNSQOPEGLLV-RAQHEGVVRG 1145
QY 1198 TALGSPVGSITKGLPSTRVPSDSAITYRGSIHTGTPA-----DVLKGTFTIRI-IG 1248
Db 1146 TA-GAVQESITRGTPASKISVETISLRLGSIQTQTPALPOAGITAEALVKGPVSRMPLE 1204
QY 1249 EDSRKLDRGREDSLPKGHVIEYGGKHVLSYEGGMSVYTQCSKEDGRSSGPPHETAAPK 1308
Db 1205 ESSPEKV-----REEAASKGHVIEYGGKHVLSYDNKNA-----REGTRSPRTAHEMSL-K 1255
QY 1309 RYDMMEGRVGAIS-----SASIEGLMGRAPRHSHPH-HLKEQHHIRGSIQTGTPR 1361
Db 1256 RSYEAVGSIKQGMRSRSPVSALEGLTCLRALP--RGSPHSDLKERTVLSGSIHQGTGR 1313
QY 1362 SYVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYQTQALGPLKLKPAHEGLVATVKEA 1421
Db 1314 ATAESFEDGL-KYPOIKRES-----PPIRAFEGAI-----TKGKP-YDG-ITTIEM 1358
QY 1422 GRSIHEIPREEL-----RHTPELPLAPRLKEGSIQTQTPLYKDYTGASGTSKKHDSVRL 1476
Db 1359 GRSIHEIPRODILTQESRKTPEVQOSTRPIEGSISQGTPIKFDNN-SQOSAIKHNVKSL 1417
QY 1477 IGSPGRTPPVHPLDMAD-ARALERACE-----ESLKSRTGTAASSGSGSTARCAPIV 1530
Db 1418 ITGPSKL--PROGLEIVPENIKVBERGYEDVKAGEPVRAHRTSVVSGSPVLRST---L 1472
QY 1531 PELGRPROSPLTYEDHGA-----PFAGHLPRGSPVTMREPTPLQEGSLSSSK-ASQDRK 1584
Db 1473 HEAPKAQLSPGLYDSSARRTPVSYQNTISRGSPMMNR-----TSDVSSSKSASHERK 1525
QY 1585 LSTSTPRE-----IAKSPHSTVPEHHPHPTISPYEHLHVRGSDVLYRSHIPLAFDPTSIPIRG 1640
Db 1526 SFLTPTQRESIPAKSPVPGVDPIVSH--SPEDPHRRSAAAGEVYRSHLPTHLDP-AMPFH 1582
QY 1641 IFLDAAAAYLPRHLAPNPTYPHLYPGYPTDALEN-ROTLINDYLTSCQMHNT 1699
Db 1583 RALDPAAYLLQRLSPSTPGTYSQIOLY-----AMENTROTILINDYITSCQMQVNL 1633
QY 1700 ATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLPVPTPGTPATAMDRLA 1759
Db 1634 -----RPDVTTRLSPREQPLGLPYPA-TRGIIDLTNMPP-TILYVPHAGGSTPPMPDRIT 1685
QY 1760 YLPTAPQPFSSR-HSSSPLSPGPHLTKPPTTSSSERDRDRDRDRDRDREREKSIILYST 1818
Db 1686 YIPGQVTFPPPPYNAASLSPGHPTHL---AAAAGAERERERERERERERERERER 1742
QY 1819 TTVEHAP---IWRPTEQSSGSGSGSGGGGGSSRPASHAHQHSPISTPTQDA-LQOR 1874
Db 1743 ERIAAPADLYLRPSEQ-----PGRPGSHGYVRSPSP-SVRTQETILQOR 1787
QY 1875 PSVLHNTGMGIITAVEPSKPTVLRSTSTSPV-----RPAATFPPTATHCPGLGTLGTYVP 1930
Db 1788 PSVFOGTNGTSVITPLDPTAQLRIMPLPSGGFPISQGLPASRYNTAADA-LAALYDAAS 1846
QY 1931 TLMEVPLLPKEAPR-----VARPERPADTGHAFAPKAPARGGLEPASPSSK 1977
Db 1847 APQMDVSKTKESKEARLEENLRSAVSEQQOQLEKNLEVERKSVQCVCCTSSALPSG 1906
QY 1978 GSEPRPLV-----PPVSGHATTARTPAKN--LAPHHASDPDPAPPASADPHRE 2024
Db 1907 KAQPHASVYVSEAGKDKGPPPKSRYEEELRTRGKTTITANFIDVITRQIASDKDARER 1966
```

```
QY 2025 KTQSKPFSIOELELRSLGLYGHGSSYSPGVEPVSPVSSPSLTHDKGLPKHLEELDKSHLEG 2084
Db 1967 GSQSDSS-----SSLSSHYETASDAIEVISPASSPAPQEKPOAYQPDVWYKANOEN 2020
QY 2085 ELRPKQPGVPLKGLGGAHLPLHLR-----PLPESQSPSSSSPLLQT--APGVKGHQRVVTLAQ 2137
Db 2021 ESTROYEGP-----LHHYRSQOESPSPQOQPLPSPSSQSEGMGQVPRTHRLITLAD 2071
QY 2138 HISEVITQDYTRHH--PQOLSAPL-----PAPLYSFCASCPCVLDLRRPPSDLYLPPPD--- 2189
Db 2072 HICQIITODFARNQVPSQASTSTFTQSPALSSTP-----VRTKTSRYSSESQST 2123
QY 2190 --HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGEDGIEBPVSPPEGMTEPG-HSRSR 2240
Db 2124 VUHPRGPRVSPENLVDRSGRSPCKSPERSHI---PSEYEPISFPQG---PAVHEROD 2177
QY 2241 VYPLYRDEQTEPSERMGSKSPCNTSQPPAFESKLTESAMVSKSKQKOEINKKLANTHN 2300
Db 2178 SMLLSQRGVDPAEQSRSDSRSPGISYILPSPFTKL-ESTSPVSKKQOEIRKLNSSGG 2236
QY 2301 EPEYNISQGTIFNMPALTGLMTYRSQAOEHASTNGLEAIRKALMGKYDOWEE- 2359
Db 2237 DSDMAAOPGTEIFNLPVAVTSGVSSRSHSPADPAS-NLGLIEDIIRKALMGSDDKVED 2295
QY 2360 -----SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGG-GKAKVSGRPSRRK 2413
Db 2296 HGVVMSPHV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKSNRK 2347
QY 2414 AKSPAPGLA--SGDRPPSVSVHSEDCNRRRTPLTNRVWEDRSPSSAGSTPPFPNPLML 2471
Db 2348 SKSPITGOSYLGTERPSSVSVHSEGDYHRQTP--GWAMEDRPSSTGSTQFPYNPLTIRM 2405
QY 2472 QGVNMASSPPPLPAGSGPL--AGPH---HAWDEEPKPLCQYETLSDE 2517
Db 2406 ----LSSPTPTQIACSAITOAAPHQONRIWEREPAPLLSAQYETLSDSD 2452

RESULT 4
NCRL_HUMAN
ID NCRL_HUMAN STANDARD; PRT: 2440 AA.
AC 075376; Q90PV5; Q90Q18;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).
GN NCOR1 OR KIAA1047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98393736; PubMed=9724795;
RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;
RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
RT transcription by interaction with the human N-COR/MSIN3/HDAC1
RT complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN [2]
RP SEQUENCE OF 782-2440 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [3]
RP SEQUENCE OF 974-2440 FROM N.A.
RX MEDLINE=99375328; PubMed=10444336;
```



RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,  
 RA Horwitz K.B., Lupski J.R., Seo H.;  
 RT "Localization of the human nuclear receptor co-repressor (hn-COR) gene  
 RT between the CMT1A and the SMS critical regions of chromosome  
 RT 17p11.2.";  
 RL Genomics 59:339-341(1999).  
 CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).  
 CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.  
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF044209; AAC33550.1; -;  
 DR EMBL; AB028970; BAA82999.1; -;  
 DR EMBL; AB019524; BAA75814.1; -;  
 DR MIM; 600849; -;  
 DR InterPro; IPR001005; -;  
 DR Pfam; PF00249; myb\_DNA-binding; 2.  
 DR PROSITE; PS00090; MYB\_3; 1.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KW Coiled coil.  
 FT DOMAIN 174 216 COILED COIL (POTENTIAL).  
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B.  
 FT DOMAIN 299 328 COILED COIL (POTENTIAL).  
 FT DNA\_BIND 437 482 SANT-A (POTENTIAL).  
 FT DNA\_BIND 625 670 SANT-B (POTENTIAL).  
 FT DOMAIN 501 557 COILED COIL (POTENTIAL).  
 FT DOMAIN 607 617 PRO-RICH.  
 FT DOMAIN 988 1816 INTERACTION WITH ETO.  
 FT DOMAIN 2055 2059 CORNR BOX OF ID1.  
 FT DOMAIN 2263 2267 CORNR BOX OF ID2.  
 FT DOMAIN 58 64 POLY-GLN.  
 FT DOMAIN 593 603 POLY-ALA.  
 FT DOMAIN 1032 1035 POLY-PRO.  
 FT DOMAIN 1707 1712 POLY-ALA.  
 FT DOMAIN 1952 1963 POLY-SER.  
 FT CONFLICT 1014 1014 L -> V (IN REF. 2).  
 FT CONFLICT 1508 1509 PP -> SS (IN REF. 2).  
 FT CONFLICT 1561 1561 W -> R (IN REF. 2).  
 FT CONFLICT 1567 1567 Q -> H (IN REF. 2).  
 SQ SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

Query Match 31.4%; Score 4147.5; DB 1; Length 2440;  
 Best Local Similarity 40.7%; Pred. No. 4.4e-128;  
 Matches 1099; Conservative 341; Mismatches 785; Indels 475; Gaps 108;

Qy 16 EPRYPHSLSPVQIARTHTDVGLEYQ--HHSRDYASHLSFGSIQPPRRRLLSEFQ 73

Db 17 OSRYPPHSPVQVTFNTRHQQEFQFAVDYRSSHLEVSQAQLQQQQQQQLRRRPSLLSEFQ 76  
 Qy 74 PCNERSQELHLRPESHYSYLPGLKSEMEFIESKRRLLELPD-----PILRSP 122  
 Db 77 PGSDRPQB--RRTSYEPFHPGSPVDHDSLESKRRLPQVSDSHFQQRVSAAVLPVHPLP 134  
 Qy 123 LLATQOPAGSDELTKDRSLITKLE--PVSPSPPHPTDPELELVPPRLSKKEELIQNDMDRVR 181  
 Db 135 ---EGLRA-SADAKDPAFGGKHEAPSPISQPCGDDONASPSKLSKEELIQSDMDRVR 190  
 Qy 182 EITWVEQOISKLKKQOOLEEAAKPEPEKPVSPPIESKHSRLVQIITYDENRRKKAFA 241  
 Db 191 ETAKVEQOILKKKQOOLEEAAKPEPEKPVSPPIESKHSRLVQIITYDENRRKKAFA 250  
 Qy 242 HRILEGLGPVPELPYNOPTDQYHENIKINQAMRKLLILYFKRRNHARKQWKQKFCOR 301  
 Db 251 HKIFEGLGPKVPELPYNOPTDQYHENIKINQAMRKLLILYFKRRNHARKQWKQKFCOR 310  
 Qy 302 YDQMLEALEKKVERENNPRRAKESKVREYKQFPEIRKQRELOQRMQSRVGQSGSL 361  
 Db 311 YDQMLEALEKKVERENNPRRAKESKVREYKQFPEIRKQRELOQRMQSRVGQSGSL 369  
 Qy 362 SMSAARSEHSEYSEIIDGLSEQENLEKQMLAVIPPMYDADQQRKIFINMGLMADPMK 421  
 Db 370 SATIARSEHSEYSEIIDGLSEQENLEKQMLAVIPPMYDADQQRKIFINMGLMADPMK 429  
 Qy 422 VYKDQVNMWSEQKETFREKFMQHPKNGFIASLERKTVACVLYLYLTKKNENYKS 481  
 Db 430 VYKDQVNMWTHKEIKDFIQHPKNGFIASLERKTVACVLYLYLTKKNENYKS 489  
 Qy 482 LVRRSY--RRRKSSQOQQOQQOQQOQQOQPMRPSQOEKDEKEKEKEAEKEEPEVEN 540  
 Db 490 LVRRNGARRGNQOIAKPSQOEKVEKEE--DKAEKTEKEEKEKDEEKEDEKEDSKEN 547  
 Qy 541 DKEDLLKEKTDGSDENDEKAEVASKGRTANSQGRKGRITRSMANSEEAIAITPQQ 600  
 Db 548 TKE---KDKIDGTA-EETEEREQATPRGKRTANSQGRKGRITRSMANSEEAIAIAA 603  
 Qy 601 S-----AELASMELENESSRWTEEMETAKGLLEHGRNWSAIARWGVSKTVSOCK 650  
 Db 604 TEEPPPLPPPEPISTETPSTRTEEMEVAKKGLVEHGRNWSAIARWGVSKTVSOCK 663  
 Qy 651 NFYFNKQRQNLDELTOQLKLMKERNARKKKKAPAAAEEAAFPVVEDEEAEAGV 710  
 Db 664 NFYFNKRRHNLNLLQKHQKTSRKPREEDVSCESVASTVSA---QEDIEDIAS-- 717  
 Qy 711 SGNEEMVEEAEALHASGNEVPREGCSGPATVNNSSDSTESIPSPHTEAAKDTGONGPKPP 770  
 Db 718 --NEENPEDSE-----VEAVK----- 732  
 Qy 771 ATLGADGPPGPPPTPRKTSRAPTEPTASEATGAPTPPAP-PSFS-APPVVPKEKE 828  
 Db 733 -----PSDSEPNATSRGNTPEPAVELE-----PTTETAPSTSPSLAVPSTKPAEDS 779  
 Qy 829 EETAAPPV-----ERGEQKPPAAEELAVDTGKAEPEVKSECTEAE 871  
 Db 780 VETQVNDISIAETAQMDVDQOEHSAGEGSCVCDPPATK--ADSDVDVEVRVPHENHASKVE 837  
 Qy 872 EGPAGKDAEAAEATGALKAEKKEGGSGRATTAKSSGA-----PQ-DSDSATSCADEV 926  
 Db 838 GDNTKERDLRA-----SEKVEPRDEDLVAAQINAQRPQSDNSDSSATSCADE- 887  
 Qy 927 DEAGGDKNLLSP---RPSLLTPTGDPANASPOK--PLDLKOLKORAAAIPTI----- 976  
 Db 888 -DVDGEPEQRQRMFMDSKPSLLNPTGSLV--SSPLKPNPLDLPLQLQHRAAVPMVMSCTP 945  
 Qy 977 -----QVTKVHEPPREDAAFTKPPAPPAPPPQNLQPESDAPQPGSSP 1019  
 Db 946 CNIPIGTPVSGVALYQRIKAMHESALLE-----EQQRQOQIDLECRSSSTP 993  
 Qy 1020 RGSRSRAPPADKKAFAEAAQKLFQDPCWTSGLFFVPPPREVIKASPHADPDSAFSTAP 1079  
 Db 1079 RGSRSRAPPADKKAFAEAAQKLFQDPCWTSGLFFVPPPREVIKASPHADPDSAFSTAP 1079



```
Db 994 CGTSKSP-----W-----EVLQPAFH-----QLITNLP 1019
QY 1080 PGHPLPLGLHDTARPYLPRPTISNPPPLISSAKHPVSLEROIGALSQG-----MSVOLH 1134
Db 1020 EGVRLP-----TTRPRPPLIPSKTIVASEK-PSFT--MGSSIQSGTGGTYLTSHNQ 1071
QY 1135 VPYSEHAKP-VGPVTMGLPLMDPKKLLAPFSGVKQEQLSRQGAQPPESLGVPTAQEAS 1193
Db 1072 ASYTOETPKPSVGSISLGLPROQESAKSATLPYIKOEFSRQNSQPEGLLV-RAHEG 1130
QY 1194 VLRGTAQSVGPGSITKGPISRPVSDSAITYRGSITHTGTPA-----DVLYKGTIYR 1245
Db 1131 VVRGTA-GAIEGSIIRGTFTSKISVIESIPSLRGSITQGTPTALPGITPEALVKGISR 1189
QY 1246 IIGEDSPSLDRGREDLPGHVIYEGKGVHLSYEGGMSVTQCSKEDGRSSGPPHETA 1305
Db 1190 MPEDSSP--EKGREBAASKGHVIEGKGHLSYDNIRNA-----REGTRSPRTAHEIS 1242
QY 1306 AKRTYDMMEGRVGRAS-----SASIEGLMGRAPPERHSPH-HLKEQHHRIGSITQG 1358
Db 1243 L-KRSYESVEGNIKQGMMSRSPVSAPEGLICRALP--RGSFSLDKERTVLSGSIQMG 1299
QY 1359 IPRSYVEAQEDYLRRAKLLKREGTPPPPPPSRDLTEAYTQALGPLKPAHEGLVAIV 1418
Db 1300 TPRATTESFEDGL-KYPKOIKRES-----PPIRAFEGAI-----TKGP-YDG-ITTI 1344
QY 1419 KEAGRSIHEIPREL-----RHTPELPLAPRLKEGSIQTGPKLYDTGASTTGSKKHDV 1473
Db 1345 KEMGRSIHELPRQDIITOESRKTPEVVOQTRPIEGSISQGTPIKFDNN-SGSAIKHNV 1403
QY 1474 RSLIGSPGRTFPVPHPLDVMAD--ARALERACYE-----ESLSRPGTASSGGSIARGAP 1527
Db 1404 KSLITGPSKLSRGMPLEIVPENIKVVERKYEDVKAGETVRSRHTSVVSSGSPVLRS- 1462
QY 1528 VIVPELKGKRSQPLTYEDHGA-----PFAGHLPRGSPVMTREPTPLRQGLSSSSRASOD 1582
Db 1463 --LHEAPKAQLSGIYDDTSARTPVSYQNTMSRSGPMNRTSDVTIP-----PNKSTNHE 1516
QY 1583 RKLTSTPRE-----IAKSPHSTYPEHHHPISPYEHLRLGSGVDLYRSHIPFLAFDPTSIP 1638
Db 1517 RKSTLTPTQRESIPAKSPVGVDPVYSH--SPFDPHRGSTAGEVYVSHLPQLDLP-AMP 1573
QY 1639 RGIPLD-AAAAYVPLPHLAPNTPYHLYPPYLRIGVDPDTAALEN-ROTIINDYITSQQMH 1696
Db 1574 FHRALDPAAAVLFQRLSFTPCYPSQYLY-----AMENTROTILNDYITSQQMQ 1624
QY 1697 HNTATAMAOADMRLGSLPRESSLALNYAAGPRGIIDLSQVPHLPVLVPTTCTPATAMD 1756
Db 1625 VNL-----RPDVARGLSPREQPLGLUPYA-TRGIIIDLTMNPP-TTLVPHPGGTSTPPMD 1676
QY 1757 RLAYLPTAQPTSSR-HSSSPSLSPGPGTTLTKPTTSSSERDRDRDRDREREKSL 1815
Db 1677 RITYIPGTQITPPRPNYSASMSGHPTHL---AAAASARERERERERERER---TA 1730
QY 1816 TSTTVEHAPIWRPGTEQSGSGSGSGGGGSSRRPASHAHQHSPISPRQDA-LOOR 1874
Db 1731 AASSDL-----YLRPGSEQ-----PGRPGSHGVYRSPSP-SVRTQETMLQQR 1771
QY 1875 PSVLHNTGMKGITAVEPSKPTVLRSTSSPV-----RPAATFPATHPLGTLGTVYP 1930
Db 1772 PSVFOGTNGTSVITPLDPAQLRIMPLPAGGFSISGGLPASRYNTAADA-LAALVDAAS 1830
QY 1931 TLMPEVLLPK-----EAPRVARPERPRAD-----TGHAFLAK 1962
Db 1831 AFQMDVSKTKESKHEARLEENLRSAAVSQOQLEKTEVEKSVOCULTSSAPSG 1890
QY 1963 PPARSGLEPASS---PSKSEPRPLVPVPSGHATTARTPAKN--LAPHIASPDPAPPAS 2017
Db 1891 KP-----QPHSSVYVEACKDKGP--PPKRYEELRTGKTTITAAFDIVIIITRQIAS 1943
QY 2018 ASDPREKTQSPFSTQIELSLRSLGVHSGSPGVEPVSPVSSPSLTHDKGLPKHLEEL 2077
Db 1944 DKDAERGSQSSDSS-----SSLSSHYETPDAIEVISPASSPAPPQEKLTQYQEVV 1997
QY 2078 DKSHLEGELRPKQPGPVKLGGEAAHPLRLPSPESOPSSSPLLQTAG-----VKGH 2129
Db 1998 KANOANDTRQTEGP-----LHHTRP-----QOESFSPQOQLPPSOAEGMGVPR 2045
QY 2130 QRVVITLAQHISEVITODYTRHH-----POOLSAPLPAPLYSFPGA--SCPVLDLRRPPSD 2182
Db 2046 HRLITLADHICQIITDFARNQVSSQTPQO---PPTSTFQNSPVALVSTPV---RTKTSN 2099
QY 2183 LYLPPPD-----HGAP-ARGSPH-----SEGKRSPEPNKTSVLGGEGDIEPVSPPEGM 2231
Db 2100 RYSPESQAQSVHHQRSGRVSLENLVDKSRGSRPGKSPERSHV---SSEYERISPPQ-- 2154
QY 2232 TEPGHSRSVAVPLLYRDGQTEPRMGSKSGNTSOPPAFFSKLTESNAMVSKKQEIIN 2291
Db 2155 VPVVEKQDLSLLSORGAEPAQRNDARSPGISVILPFFTKL-ENTSPMVSKKQEIF 2213
QY 2292 KKLTHNRNEPEYINISOPGTEIFNMPAITGTGLMTYRSQAVQEHASTNMGLEAIIRKALM 2351
Db 2214 RKLNSGGGSDMAAOPGTEIFNLPVATTSGSVSRGHSFADPAS-NLGLEDIIRKALM 2272
QY 2352 GKYDQHEE-----SPPLSANAENPLNASLPAAMPITAAAGRSODHTLTSPGGGG--KA 2403
Db 2273 GSFDDKVEDHGVVMSQPMGV---VPGTANTSV-----VTSGETRREEGDPSHSGGVCKP 2324
QY 2404 KVSGRSPSSKAKSPAP--GLASGDRPPSVSVHSEGDCHNRRTPLTNRVWEDRPSASGTP 2461
Db 2325 KLIKSNRSKSKSPIGQYLGTERFSSVSVHSEGDYHRQTP--CWAWEDRPSFGSTQ 2382
QY 2462 FYPNPLIMRLQAGVMASPPPPGLPAGSG-PLAGPH---HAWDEEPKPLCSQYETLSDSE 2517
Db 2383 FYPNPLTMRM--LSSTPTPTIACAPSAVNOAAPHQONRIWEREPAPLLSAQYETLSDS 2439
RESULT 5
NCRL_RAT
ID NCRL_RAT STANDARD; PRG: 533 AA.
AC Q9WUB5; O70463;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-PRESSOR 1 (N-COR) (N-COR) (FRAGMENT).
GN NCOR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99371771; PubMed=10441327;
RA Boutell J.M., Thomas P., Neal J.W., Weston V.J., Duce J., Harper P.S.,
RA Jones A.L.;
RT "Aberrant interactions of transcriptional repressor proteins with the
RT Huntington's disease gene product, huntingtin.";
RL Hum. Mol. Genet. 8:1647-1655(1999).
RN [2]
RP SEQUENCE OF 476-528 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=99421707; PubMed=10491148;
RA Schuler M.J., Buehler S., Pette D.;
RT "Effects of contractile activity and hypothyroidism on nuclear hormone
RT receptor mRNA isoforms in rat skeletal muscle.";
RL Eur. J. Biochem. 264:982-988(1999).
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
```











Db	1573	KPEPGQAESRDTGTTEALTPHINNLHTATSRKSYRPTSMPEWMEPLSPEDVAGTMS	1632
Qy	2038	LRSLYGHSSYSPEGVEPVSSVSLTHDKLG-----PKH-----LEELDKSHLEG	2084
Db	1633	QSDSGVDLSGDSQSVSSGPCSORSSP---DGSLKAAEGPKPRPGSGSPLNAVCEGPPG	1688
Qy	2085	ELRPKQPGVKLGGEAAHPLHPLPESOPSSPLLQAPGVKGHQRVVTAQIHSEVIT	2144
Db	1689	SEPRRRPPAPHDGDKRELPRQPLP-----PGPIGTER-----SQR-----T	1726
Qy	2145	QDTRHHPOOLSAPLPAFLYSFPGASCPLVLDLRR-----PPS-DL	2183
Db	1727	DRGTEGPIRPS-HRGCPPVQF-GTSDKDSDLRVVGDLSKAELKTASVTEALPVRSDW	1784
Qy	2184	YLPPPDHGAARGSPHS-----EGGKRSPEPN-----KTSVLGGGEDGI-	2222
Db	1785	ELLP---SAAASAEPQSKNLDGCHGVPEPSSGQRLYPEVFGSAGPSSSQISGGSHGLS	1841
Qy	2223	-----EPVSPPEGWTEPGHRSNAV-YPLLYRUGEQTEPSRMGSKS	2261
Db	1842	ITSQWNLRLPGTSLPHYSQQLYLPFGAPPALLSGVALKGQLDFSTWQATLKGKL	1901
Qy	2262	PGNTSOPPAFF-----SKLTESNAVMKSKKOEINKLNTNHNPEYNIHQGTET	2313
Db	1902	AGGLVLPSPSFLYSPAFCPSPLDTSLLQVRQ-----DLFSP-SDF	1941
Qy	2314	FNMPATIG-TGLMTYRSQAVQEHASTNMGLEALIRKALMGKYDQWESPLSANAENPL	2371
Db	1942	YSTPLQPGGSGFLPSGAPAQO-----MLPMWDSQLPVVNGSLPPAPP---PAPPL	1992
Qy	2372	NASASILAAMPITAAADGRSDHTLTSPGGGKAKVSGRPSRKAKSPAPGLASGDRPPSVS	2431
Db	1993	SLLPVGPALQPPSLA-----VRPPPAATRVLPSA-----RPPAS	2029
Qy	2432	SVUSE-----GDCNRRTPLTNRVWEDRPSASGTPPYNPLI--M	2469
Db	2030	LGRAELHPVELKPFODYQKLSSNLGPGSSRTPPTGTSFGSLNRLKATPSTYSGVFTQ	2089
Qy	2470	RLQAGVNASP-----PPPLPAGSGPLAGPHAMDEEP	2502
Db	2090	RVDLYQOASPPDALRWIPKWPWERTGPPREGPSRR-AEEP	2128
RESULT	7		
MAPA	RAT		
ID	MAPA_RAT	STANDARD;	PRT; 2774 AA.
AC	P34926;		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DE	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].		
GN	MAP1A.		
OS	Rattus norvegicus (rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid:10116;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=92355629; PubMed=1379599;		
RA	Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;		
RT	"Microtubule-associated proteins 1A and LC2. Two proteins encoded in		
RT	one messenger RNA."		
RL	J. Biol. Chem. 267:16561-16566(1992).		
CC	-!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS		
CC	CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.		
CC	-!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE		
CC	WITH MAP1A AND MAP1B PROTEINS.		
CC	-!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.		
CC	-!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT		
CC	APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE		
CC	THEIR MORPHOLOGY.		











Db 1759 -----QESSPOKLEVER-----WLAESPVLGPEEEDKLTSPFETISPPA 1800  
QY 713 NEEEMVEEAALHAGSNEVPRCEGSGPATVNNSSDTESTPSPTHEAAKDTGQNGKPPAT 772  
Db 1801 SPPEWVGQ-RVPSAGQESP-----IDPKLMPHMKNEPTTPSWLADIPWPVKDRPLPPAP 1856  
QY 773 LGADGPPGPTTPPTTRAPTEPTASATGAPPPP-----APPSAP 818  
Db 1857 L---SPAGPPTP-----APESHTPAPFSWCTPEYDSVVAQVGAEEGPGYSPGLK 1907  
QY 819 PVPVKEKEEETAAPVVEEGEOKPPAAELAVDTGKAE-EPVKSECTEAEAE----- 872  
Db 1908 DYRKVXGEREEGRAPDKSHXKVPKARKSHATTEPEQTEPEQREPTPYPDERSFOY 1967  
QY 873 -----GPAKDXDAEAAETAGCAKAEKGGSGRATTAK-----SSGAPQSD 916  
Db 1968 ADIYEQMMLTGLGPACPTREPLPGLAGWPPCLSKYKAAAGNRTSAEKLSPISPKSLQ 2027  
QY 917 S-----SATCSADEVDEAEGDKNRLSPRLTPTGDPRA---NASPOKPLD---L 963  
Db 2028 SDTPTFSYAALAGPTVPPRXEPGP-----SMEPSLTPPAVPPRAPILSKGSPPLNGNLL 2082  
QY 964 KOLKORAAIP-----PIQTKVHEPPREDA--APTKPAPPPP--PQNLQPE 1009  
Db 2083 SCSPPDRSPSPKESGRSHWDDSTSELEKGAQAEKAEQSPSPHPPTPMGSPTLWPET 2142  
QY 1010 DAPQOPG-SSPRGKSRSPAPPADKFAFAEAKL-----PGDP--PCWTSGLFP- 1055  
Db 2143 EAHVSPPLXSHUGXAR---PSLDFPASAGFSSLEXAPQXPSPAEPRAPC--GSLAF 2197  
QY 1056 -----PVPPR-----EVIKAS-----PHAPDPSAFSPAGPHPLPLGHDHAR 1093  
Db 2198 GDRALAPAGPTTRYDEYLEVTRKASLSDSLPQLPSPSS-----PCXPLLSNLPRAS 2252  
QY 1094 PVLPRPTISNPPPLISSAKH---PSV--LERQIGAIQSGMSVQLHVYSEHAKAPGV 1148  
Db 2253 PALSGSSEATTPVISSVAERFSPLEAAEQESGELDPGEPAHXLWDLTPLSPAPPA 2312  
QY 1149 TMGL---PLPMDP-----KKLAPSGVYKQELSPRG--QAGPPESL 1184  
Db 2313 SLDLALAPSLPGMDGILPCHLECESEATEKPSFQ-VPEOCAANGPTETSPNPPX 2371  
QY 1185 GVPTA---QASVLRGTALGSPVGSITKIPSTRVPSDSAITYRGTHGTPADVLYKG 1241  
Db 2372 PAPAKAENEAAAXPAWEGAWPEGAERSSRDXTLSPEQPV----- 2413  
QY 1242 TITRIIGEDSPRLDRGDSLPKCHVIYEGKKHVLVYEGGMSVTCQSKEDGRSSGPP 1301  
Db 2414 -----CPXG-----GSGGPP 2423  
QY 1302 HETAAPKRTYDMGVRGVAISSASIE---GLMGRAIPPERHSPHLKEOHIRGSITQG 1358  
Db 2424 -----SSASPEVEAGQGCXTERPH-----RGELSPS 2451  
QY 1359 -----IPRSYVEAQEDYLRREAKLKR-----EGTPPPPPPSRDLTEAYKTQALGPL 1405  
Db 2452 FLNPLPSPIDD--RDLSTEEVRLVGRGRRVGGGPGTGXPVPTDETPPTSADSG-- 2507  
QY 1406 KLKPAHEGLVATVKAGRSIHEIPREELRHPELPLAPRLKESITQCTPLKYDTGAST 1465  
Db 2508 -----SSQSDSDVPPE---TECPSTIAEAALDDEDEDFLPVCKXGV 2548  
QY 1466 TGSKK---HDVRSLLIGSGRTFPVHPPLDVNMADARAL-----ERACYEESLSRGT 1514  
Db 2549 SETHPRGHDPPLPQDPRSPB-RPDCVMADPEGLSSSGRXXERLXXEKVQVRGV- 2606  
QY 1515 ASSSGGSARGAPVIVPELGRPQSPFLYEDHCAFAGHLPGRSPVTWREPTPRLEQSL 1574  
Db 2607 -----RRAP-----GDKK--PV-----SPXRRLXLRGK---RSPTPGKSDXR 2639  
QY 1575 SSSKASQDRKLTST--TPREIAKSPHSTVPEHHPHPI--SPYEHLLRGVSGVDLYRSHIPL 1630

Db 2640 VSRXPFRSRSTXSVQTPAE-EKDGHSMPKGLVNGLKAGXPALSKGSGGAPVY---VDL 2695  
QY 1631 AFDPTS-IPRGIPLO-----AAAYYLPRHLAPNPYPHYLPYLRGYDPTAAL----- 1679  
Db 2696 AYIPNHSKGTADLOFFRRVRASYIVVSGNDP-----ANGEPSRAVLDALEGG 2743  
QY 1680 -----ENRQ---TIINDYITS-----QQMH 1696  
Db 2744 KAQWGENLQVKTLIPTHDTEVTREWYQOOTH 2774  
RESULT 9  
TRX2\_HUMAN  
ID TRX2\_HUMAN STANDARD; PRT: 2715 AA.  
AC Q9UMN6: Q9UK25; Q95836; Q9Y669; Q9Y668; O15022;  
DT 01-OCT-2000 (Rel. 40; Created)  
DT 01-OCT-2000 (Rel. 40; Last sequence update)  
DT 01-OCT-2000 (Rel. 40; Last annotation update)  
DE TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN).  
GN TRX2 OR HRX2 OR MLL2 OR KIAA0340.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,  
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,  
RA Lamerdin J., Chambon P., Lossos R., Stewart A., Aasland R.;  
RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin  
RT regulators which contain PHD fingers and SET domains."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,  
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Ganes J., Danganan L.,  
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
RA Carrano A.V.;  
RT "Sequence analysis of a 1 Mb region in human 19q13.1";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).  
RC TISSUE=Brain;  
RX MEDLINE=9734984; PubMed=9205841;  
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VII.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro."  
RL DNA Res. 4:141-150(1997).  
RN [4]  
RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).  
RC TISSUE=Testis, and Leukocyte;  
RX MEDLINE=20105772; PubMed=10637508;  
RA Huntsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,  
RA Wiedemann L.M., Aparicio S., Caldas C.;  
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps  
RT to 19q13.1 and is amplified in solid tumor cell lines";  
RL Oncogene 18:7975-7984(1999).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
RC TISSUE=Placenta, and Bone marrow;  
RX MEDLINE=99339983; PubMed=10409430;  
RA Fitzgerald K.T., Diaz M.O.;  
RT "MLL2: A new mammalian member of the trx/MLL family of genes.";  
RL Genomics 59:187-192(1999).  
CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.  
CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,  
CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL







```
Db 1170 RVRVDFKEDCDLENLWMLGGLSVLTAVPGPPMVCILLC-ASKGLHELVCQCCDPFFHF 1228
QY 1442 ---APRPLKEGSIQTGTPKLYDT-----GASTGSK-----1469
Db 1229 CLEEAERPL-----PQHHTWCCRCRCKFCHVCGKRGKSKHLLCECERHAYHPAC 1279
QY 1470 -----KHDV-----RSLGSGPRTFPPVHPLDVMADARALE-----1500
Db 1280 LGPSYPTATRRKRHHWICSACVRCCKSGCATPGKNDWESGDYSLCPRCTQLYEKNYCP 1339
QY 1501 ---RACYE-----ESLSRPGTASSGSGSIARGA---1526
Db 1340 ICTRCYEDNDYESKMQCAQCDHWYHAKCEGLSDEDYETLSGLPDSVLYTCGPCAGAAQP 1399
QY 1527 -----PVLVPELKGPRQSPPLTYEDHGAPFAGHLPRGSPVMTREPTPRQOE 1571
Db 1400 RWREALSALQGLRQVLOGLSSKVVGGLLCTQCGDGGKQLHPG-PCGLQAVSORFED 1458
QY 1572 GSLSSKASQDRKLTSTPREIAKSPHSTVPE-----HHPHPISPYEBL---LRGV---1618
Db 1459 GHY-----KSVHSEFMDVMVGILMRHSEGETPDRRAGGQMKGLLLK 1499
QY 1619 ---SGVDLYRSHIPAP-DPTSIPIRGIPIDAAAYLLPRHLAPNPTYPHLY-----1665
Db 1500 LLESAGFWDHDPKRYRRSTRPLNGV-----LPNAVLP-PSLDHVYAQWRQBPPE 1549
QY 1666 -----PPYLIRGYP-----DTAAL-----ENR 1682
Db 1550 TPESQOPP-----GDSAAAFQKQDPAFASHLEDPRCALCLKYGDADSKAERLLYIGONE 1605
QY 1683 QTIINDYITSQOMH-----NTATAMAQRADM-----1709
Db 1606 WTHVNCALWAEVFEENDGSLKNVHAAVARGMRCCLCKPGATVGCCLSKLSNHFEM 1665
QY 1710 -----1715
Db 1666 CARASYCIFQDDKVFCKQKHTDLLDKEIVNPDGFDVLRVYVDFEGINFRKRLTGLEP 1725
QY 1716 RESSALNVAAGPRIIDLSQVPHL-----PV-----LV 1744
Db 1726 D-----AINVLGSIIDSGLTSLDSCEGRFPFIQYQCSRLYWTVDARRCWTACRILL 1781
QY 1745 PPTPGTATAMDRLAYLPTAPFPSSRHSSPLS---PGSPHLLTKPTTSSSERERDRDR 1802
Db 1782 EYRWMPG---REPAHLEAAEENQIVHSPAPSEPPG-----1817
QY 1803 ERDRERKESILTTTIVVEHAPIWR---PGTEQSSGSGSSGGGSSSRPASHAHQH 1860
Db 1818 ---EDPPLDLDVLPFAPERHSPIONLDPPLRPDSGSA-----1852
QY 1861 SPISPRTOALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVRPAATFPPTHCP 1920
Db 1853 PPAPRS-----FSGARKVNYSPSRR-----P 1876
QY 1921 LGGTLGQVYPTLMEVLLPVPKAPRVARPERPRADTGHAFKAPPAR-----SGLEPASSPS 1976
Db 1877 LGGVSGFLPSPGSPSSLTHIPTVGDPDF-----APRRSRPSPLAPRPPS 1926
QY 1977 KGEPRPL-----VPPVSGHATTARTPAKNLAPHASDPDPAP-----2014
Db 1927 RWASP-PLKTSPLQRVPPTSVVTALTSTGELAPPGAPSPPPPPDLGPDFEDMEVVS 1985
QY 2015 -PASADPHREKTSQPSIOBELRSYGHGSSVSPGVE-----PVSPV 2059
Db 1986 LSAOLDFAASLLGTEFP--QE-EIVAAGAMGSSGGGFGDSEESSSTSRVIHFPVTIV 2042
QY 2060 SSPSL--THDKGLPKHLELD-----KSHLEGERLPK-----QPGVKILGGAHL 2104
Db 2043 SAPGLAPSATPGAPR-IEQLDGVDDGTDSEAEAVQPRGQGTTPPSGPGVVRAGVLGAAGD 2101
QY 2105 HURLPESQSSSPLLQATPAVGKQHVVTVAQHTSEVITODYTRHHFQQLSAPLAPLY 2164
||| |||
```

```
Db 2102 RARP-PEDLP-----SEIV--DFVL---KNLGGP-----2124
QY 2165 SPPGASCPVLDLRRPPSDLYLPPDPHGAP--ARGSPHSEGGKRS-EPNKTSLVGGEDG 2221
Db 2135 GDGGAG-----PREESLPP-----APPLANGSOPSOGLTASPADPTRTFAWLPAGP 2171
QY 2222 IEPVS-----PPEGMTEP---GHSRSAYVPLLYRDGEQTEPSRMGSKSPGNTSOPPAFF 2272
Db 2172 VRVLSIGPAPEPPKATSKIIILVNLGQVFKMAGEGPPVP-----VKQPLP---2220
QY 2273 SKLTESNAMVKSQKQEKINKLNTNRPENIYINPQTEIFNMPAITGTGLMTRSQAV 2332
Db 2221 -----PPTISPTAPTSTWTLPPGLLGLVPV-----2246
QY 2333 QEHASTNMGLEALIRKA-----LMGKYDQWEESQPLSA-----NAFNPLNASASLPA 2379
Db 2247 -----GVVRPAPPPPPPLTLVSSGPASPPROAIRKRVSTFSGRSPAPPY 2295
QY 2380 AMPITAADGR-SDHTLTSPGGGKAKVSGRSPSRKAKSP-APGLASGDRPPSVSVHSEG 2437
Db 2296 KAPRLDEDEGEASEDTPOVPLG-----SGGFSKVRMKTPTVRGVLDRPGEAGEESPG 2350
QY 2438 DCNRRTPLTNRWEDRPSAGSTPPFYNPLMRLO-----AGVNAS-----PPPGGLPAGSG 2489
Db 2351 PLQERSPLL-PLPEDGPPQVDPGP---PDLLLESQHHHYSGEASSEEPSPDDKENOA 2406
QY 2490 P-LAGPH 2495
Db 2407 PARTGPH 2413

RESULT 10
MAPB_MOUSE
ID MAPB_MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2) (MAP1(X)) [CONTAINS: MAP1
DE LIGHT CHAIN LC1].
GN MAP1B OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT and tau.";
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
J. Cell Biol. 109:3367-3376(1989).
CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
CC KKEE AND KKEI/V. REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
CC BOTH IN VITRO AND IN VIVO.
CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -!- SIMILARITY: TO NEURAXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X51396; CAA35761.1; -  
DR PIR; S07549; QRMSP1  
DR MGD; MGI:97179; Mtap5.  
DR InterPro; IPR000102; -  
DR Pfam; PF00414; MAP1B\_neuraxin; 10.  
DR PROSITE; PS00230; MAP1B\_NEURAXIN; 7.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2464  
FT DOMAIN 589 787  
FT LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
FT KREE AND KREI/V REPEATS).  
FT 12 X 17 AA TANDEM REPEATS.  
FT 1.  
FT 2.  
FT 3.  
FT 4.  
FT 5.  
FT 6.  
FT 7.  
FT 8.  
FT 9.  
FT 10.  
FT 11.  
FT 12.  
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 3.2%; Score 424.5; DB 1; Length 2464;

Best Local Similarity 19.3%; Pred. No. 3.4e-07;

Matches 521; Conservative 314; Mismatches 996; Indels 869; Gaps 114;

QY 68 LLSFQPGNRSQELHRLPESHLYPELG-----KSEMEFTESKRPRLLELPDPLL 118  
DB 174 LLSSTHPANKASLT-----FCPEGDWKNLDRHNLQDFINIKLNSAIIIP----- 221  
QY 119 RPSLLATQAPAGSBDLTKDRSLTKLEPVSPSPHTDPELELYPP-----RLSKE-- 170  
DB 222 -----EMGLSEFTYLS-----ESVEVPSPFDI-----LEPPTSGGFLKSKPCC 262  
QY 171 -----ELIQNMDRVDRITWVEQOI----- 190  
DB 263 YIFPGRGDSALFVNGFMNLINGSERKSCFWKLTIRHLDRVD-----SILLTHIGDDNLPG 319  
QY 191 --SKLKKKOOOLEEAAKPPPEPEKVPSPPIESKHSRLVQIIYDENRKKAAEAHRIEGL 248  
DB 320 INSLQRTAELEEE-----RSQGSTNSNDWMKNLI 350  
QY 249 GPQVELPLYNQSDTRQYHENIKINQAMRKKLILFKRRNHARKQWKOKFCORYQOLMEA 308  
DB 351 SDIAGVVLNPNELKDPENIKMRKSIEEACFTL-----QYLNKLSMKPEPLFRS 401  
QY 309 LEKKVE-----RIENPRRRAKESKVREYVEKQFPEIRKOR-ELQERMOSRVQY 356  
DB 402 VQNTTEPVILFQKMGVGLKEMVLPNPKVSSKEMQFMQWGTCTNKDKAELIILPQGEVDI 461  
QY 357 RGSGL-SMSAARSEHVS-----EIIDGLSEQENLEKQMRQLAV----- 394  
DB 462 PISYLTSSVSLIWHPANPAEKIIIRVLPGNSTQYNILEGLEKHLKHLDFLQPLATQKDL 521  
QY 395 -----IPMLYDADQQRKIFINMGLMDPMKVYKDRQVMNMSQEQKE-TFRKFMQHP 448  
DB 522 TQQVTPPVQVYKLRQ---ADRSRESLPATPKPVASKSVRKESKEETPEVTNTKSQVETP 578  
QY 449 KNFGLIASFLERTVAECVLYLYLKNNENYKSLVRRSVRRGKSQOQOQO-QOQOQOQO 506  
DB 579 K-----VESKEV-----LVKKDKPVKTESKPSVTEKESKEEQSPVKAEEVAK 623

QY 507 QOOQPMRSPSQBEKDEKE-KEKEAEKEBEKPEVENDKEDLLKEKTDITDTSGEDNDEKAVA 565  
DB 624 QATESKPKVTKDKVVKELKTKLEKKEKPK-KEVVKEDKTKPLKKDKPKRKEEVK 679  
QY 566 SKGRITANSQGR------KGRITRSMANEANSEEAITPQOSAEALSMELN 610  
DB 680 KEIKKEIKKEERKELKVKETPLKDAKKEVKEKKEKKEK-EPKKEIKKISKDIK 738  
QY 611 ESSRWTEEMETAKKGLLEHGRNWSAIAIMVSKVTSOCKNFYFNKKRONDEILOQUK 670  
DB 739 KST-----FQSDTKP-----SALKPKVAKKEESTKK-----EPLAAGK 772  
QY 671 LKMEKERNARRKKK-APAAASEEAAFPVVEDEMEASGVSGNEEMVEEAALHA 726  
DB 773 LKDKKVKVVIKKEGKTTAAATAVGTAATAAATAAAGIAASG-----PVKELEA--- 822  
QY 727 SGNEVPRGECGPATVNNSSDTEIPSPHTEAAKDTGONGKPPATLGCADGPPGPTTPP 786  
DB 823 -----ERSILMSPEDL--TKDFEELKAEIDVAKDI-----KPQLELIED----- 860  
QY 787 RRTSRAPIEPTPASEATCAPTPPPAPSPSPAPPPVVPVKEEKEEETAAAPPVE-----BGE 841  
DB 861 -----EELKETQPGAY-----VIQETEVSKGSASPDGIGITTTGE 899  
QY 842 EOKPAAEELAVDTGKAEPPVKSECTEAEAGPKAGKDAEAAEATAGALKAEKKEG--- 898  
DB 900 GECEOTPEEL-----EPVEKQGVDDIEKFEDESGAGFESSETGDYEEKAEETAEAP 951  
QY 899 ---GSGRATTAKSGAQODSSSATCSADEVDEABEGDKNRLSPRSLTPTGTGPRANA 955  
DB 952 EBDGEDNASGASKHSPTEDDESAAEADV-----HLKKEKRESV--SGDDRAEE 999  
QY 956 SPQKPLDLKQLKQRAAAIPIQVTKVHEPPREDAAFTKAPAPPAPPPQNLQPSD----- 1010  
DB 1000 DMDDVLEGEAFQS-----EEGEEDKADAREEGYEPDKTAEDYVMAVADKAAEA 1052  
QY 1011 -----APOQPG--SSPRGKRSRPAPD-----ADKEAFAEAAOKLPGDPPC 1048  
DB 1053 GVTERQYGLTSAKQPGIQSP---SREPASSIHDETLPGGSESEATASDENREDQDEE 1109  
QY 1049 WTSGLPFPVPREVITKASHAPDPSAFYAPGHPPLPLGLHDTARPVLP-----RPTTSN 1104  
DB 1110 FTATSGYQSTIEI-----SSEPTMDMSTPRDVMSETNNEETESP 1152  
QY 1105 PPPLTSSAKHPSVLERQ-----ICAIQSGMSVQLHVPVSEHAKAPVGTMLPLP 1155  
DB 1153 SQEFNITKYESSLYSQEYKPAVASFNLSEGGSKTDATDKGDNYSAS-----TSPSS 1208  
QY 1156 MDPKKLA-----PFGSVKQEOQLSPRGQAGPPESLSGVPYTAQEAASV 1194  
DB 1209 MEEDKFSKALRDAYCSEKELKASAEIDIKDVSDERLSP--AKSPSLSPSP-----SP 1261  
QY 1195 LRGTALGVPGSGITKGPSTRVPSDSAITYRGSTHGTTPADVLYKGTITRIIGEDSPSR 1254  
DB 1262 IETPLGE-----RSVNFSLTPNEIKVSAEARSVSP-----GVTQAVVEHCASP 1308  
QY 1255 LDRGREDSLPGHVIYEGKGHVLSYEGGMSVTCQSKEDGRSSSGP-----PHETAAPKRT 1310  
DB 1309 BEKTLVYVSPQSV--TGSAGHTPYQ-----SPTDEKSSHLPTVESENAQAVP--- 1355  
QY 1311 YDMBGRVGRAISSASIEGLMGRAP-----PERSHPHLKEQHHRGSIQTGIPRSVVE 1365  
DB 1356 -----VSFESEAKDENERSLSPMDPEVPDSESPVE-KVLSPLRPLILGSESPY-- 1405  
QY 1366 AQEDYLREAKLLK-----EGTPPPPPPSRDLTEAYKTAQALGPLKPAHEG 1413  
DB 1406 --EDLSADSKVGLRRSESPFEGKNGKQFPDRESVPVSDLT-----STGLYQDKQEKSTG 1459  
QY 1414 LVATVYKAGRSITHEIPREELRHTPELPLAPRLKSGITQGTPLKYDTGASTGSKKHVD 1473  
DB 1460 FIPKEDGPEKKTSDVETMSSQSALADERKL--GGDV---SPTQID--VSQFGSKEDT 1513  
QY 1474 RSLIGSPGRTFPVPHPLDMADARALERACYEESLKSRRGTASSSGGSIGAPVIVPEL 1533



















RP REVIEW ON OI VARIANTS.  
RX MEDLINE=91374476; PubMed=1895312;  
RA Byers P.H., Wallis G.A., Walling M.C.;  
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";  
RL J. Med. Genet. 28:433-442(1991).  
[13]  
RP REVIEW ON OI VARIANTS.  
RX MEDLINE=97169389; PubMed=9016532;  
RA Dalgleish R.;  
RT "The human type I collagen mutation database.";  
RL Nucleic Acids Res. 25:181-187(1997).  
[14]  
RP VARIANT OI-II CYS-1166.  
RX MEDLINE=86287390; PubMed=3016737;  
RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;  
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human pro alpha 1(I) collagen allele.";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).  
[15]  
RP VARIANT OI-II ARG-569.  
RX MEDLINE=87222295; PubMed=3108247;  
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;  
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of arginine for glycine at residue 391 of the alpha 1(I) chain of type I collagen.";  
RL J. Biol. Chem. 262:7021-7027(1987).  
[16]  
RP VARIANT OI-II CYS-926.  
RX MEDLINE=88033031; PubMed=3667599;  
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;  
RT "A point mutation in a type I procollagen gene converts glycine 748 of the alpha 1 chain to cysteine and destabilizes the triple helix in a lethal variant of osteogenesis imperfecta.";  
RL J. Biol. Chem. 262:14737-14744(1987).  
[17]  
RP VARIANT OI-II ARG-842.  
RX MEDLINE=88298828; PubMed=3403550;  
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;  
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I) chain in lethal perinatal osteogenesis imperfecta. Demonstration of the peptide defect by in vitro expression of the mutant cDNA.";  
RL J. Biol. Chem. 263:11627-11630(1988).  
[18]  
RP VARIANT OI CYS-1195.  
RX MEDLINE=89218628; PubMed=3244312;  
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;  
RT "A cysteine for glycine substitution at position 1017 in an alpha 1(I) chain of type I collagen in a patient with mild dominantly inherited osteogenesis imperfecta.";  
RL Mol. Biol. Med. 5:197-207(1988).  
[19]  
RP VARIANT OI-II VAL-434.  
RX MEDLINE=89255493; PubMed=2470760;  
RA Patterson E., Smiley E., Bonadio J.;  
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta mutation.";  
RL J. Biol. Chem. 264:10083-10087(1989).  
[20]  
RP VARIANT OI-IV SER-1010.  
RX MEDLINE=89308591; PubMed=2745420;  
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;  
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";  
RL J. Biol. Chem. 264:11893-11900(1989).  
[21]  
RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.  
RX MEDLINE=89380165; PubMed=2777764;  
RA Lamande S.R., Dahl H.H., Cole W.G., Bateman J.F.;  
RT "Characterization of point mutations in the collagen COL1A1 and COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";  
RL J. Biol. Chem. 264:15809-15812(1989).  
[22]

RP VARIANT OI SER-1022.  
RX MEDLINE=90062068; PubMed=25111192;  
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;  
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe variant of osteogenesis imperfecta minimally destabilizes the triple helix of type I procollagen. The effects of glycine substitutions on thermal stability are either position of amino acid specific.";  
RL J. Biol. Chem. 264:19694-19699(1989).  
[23]  
RP VARIANT OI-II CYS-1082.  
RX MEDLINE=89109573; PubMed=29130553;  
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;  
RT "A lethal variant of osteogenesis imperfecta has a single base mutation that substitutes cysteine for glycine 904 of the alpha 1(I) chain of type I procollagen. The asymptomatic mother has an unidentified mutation producing an overmodified and unstable type I procollagen.";  
RL J. Clin. Invest. 83:574-584(1989).  
[24]  
RP VARIANT OI CYS-272; CYS-704 AND CYS-896.  
RX MEDLINE=90009313; PubMed=2794057;  
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weiss M.A., Weiss L., Graham J.M., Byers P.H.;  
RT "Osteogenesis imperfecta. The position of substitution for glycine by cysteine in the triple helical domain of the pro alpha 1(I) chains of type I collagen determines the clinical phenotype.";  
RL J. Clin. Invest. 84:1206-1214(1989).  
[25]  
RP VARIANT OI-II CYS-422.  
Query Match 3.0%; Score 400.5; DB 1; Length 1464;  
Best Local Similarity 21.5%; Pred. No. 1.3e-06;  
Matches 380; Conservative 102; Mismatches 642; Indels 647; Gaps 83;  
QY 722 EALHAGNEVPRGECGPGATVNNSSDTEIPSPHTAAK-DTGQNGPKPPA-TLGDAGPP 779  
DB 77 ETKNCPGAEVPEGEC-PVCPDGSESPTDQETTVGEGKGTGPRGPRGPPGRDGIP 135  
QY 780 -----PGPTTPRRTSRAPETTPASEANGATPP-----P 810  
DB 136 GQPLPLPGPPPP-----GPPGPPGLGNGFAPQLSYGDEKSTGGISVP 178  
QY 811 APPPSAP-----PPVVKKEKEEETAAAPPVEEGE-----EOKPAAAEALAYDTCK 857  
DB 179 GPMGSPGPRGLPGPGAGPQGFQ-----GPPGEPGPGASGPMGPRGPPGPKNGKDDGE 234  
QY 858 AEPVKSECTEAEAGCPAKGDAEAAEAAGALKAEKKEGSGRATTAKSSCAPQSDS 917  
DB 235 AGKPGR-----PCEKPGPPGQARGLPFTA--GLPGMKGHRGFGSLDGAKGDAGP----- 282  
QY 918 SATCSADEVDEAEGDKNLLSP-----RPSLLTPTGDRANASPKPQLDLKQLKQRAAI 973  
DB 283 -----AGPKGFGSPGNGCAPQMGPRGLPGERGPGAP-----GPAGAR 322  
QY 974 PFIQVTKVHEPPREDAAATKPA-PPAPPPQNLPQSDAPQPGSS--PRGKSRSPAPP- 1029  
DB 323 GNDGATGAAGPP-----GPTGAPGPPGPGAVGAKGEA-GPQGRGSEGGPQGVGEPGPG 377  
QY 1030 -----ADKEAFAAEOKLPGDPCCWTSGLP-FPVPPREVIKASHPADPSAFS 1076  
DB 378 PAGAGPAGNPGADGQPGAKGANGAPG-----IAGAPGPPGARGSPGPGPGPKGN 432  
QY 1077 YAPPGHPLPLGLHDTARPVLPRTTISNPPPLTSSAKHPSVLERQITGAIISQGNVQLHVP 1136  
DB 433 SEEPAGPCKG-DTGAKGEPGVGVQGP-----GPAGEGKRG----- 470  
QY 1137 YSEHAKAPVGVPTMGLPLPMDPKKLPAPFSVGVKQEQLSRQAGPPESLGVPTAQEASVLR 1196  
DB 471 -----ARGECPPT--GLP-----GPPGERGGPGSGRPPGADGVAGPK 505  
QY 1197 GTA--LGSVPGGSITKGP-STRVPSDSAITVRSITHTGTPADVLTKGTTIRIGEDSPS 1253  
DB 506 GPAGERGS-PGPAGPKGSPGEAGRPGKGLT-GSPGSPGPDGK-----TGPGPGA 559











QY 2023 REKTOSKPFISIOELRSLGICGSSYSPGCVSPVSSPSLTHDKGLPKHLEELD KSHL 2082  
Db 1219 TQCLLQCPWPLNEAQVA-----SVVKVLTTEL-----L 1246  
QY 2083 EGELRPKOPGVKLGGEAAHPLRLPESQSSPLLQATPGVKGHORVVTLAQHISEV 2142  
Db 1247 EQE-RKKVVDTTKESSRKGWESKRKLSDQPA-----RTPRSKKKKL----- 1290  
QY 2143 IQQDYTRHHHPQOLSAPLAPLYSFFGASCPVLDRPPSDLYLPPDHGAPARGSPHSEG 2202  
Db 1291 -----GEG 1295  
QY 2203 GKRSPBNKTSVLGGEDGIEPVSPPEGMTEPGHRSASVYPLLYRDEGEQTEPSR-----M 2257  
Db 1296 GASVSPKTSSTSGKAKRDKAS---GDVKKKKGSGS-----LGSOGAKDEPEEELQKGM 1348  
QY 2258 GSKSPGNTSQPPAFKTSKLTESNAMYKSKQKQIN---KKLTHNRNEP 2302  
Db 1349 GTVEGGDQSNPKSKKKEKKDKRKKDKKKKKAKKASTKDSSEP 1395

## RESULT 15

CA11\_CANFA  
ID CA11\_CANFA STANDARD; PRT; 1460 AA.  
AC Q9X5J7:  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.  
GN COL1A1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_taxid=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;  
RT "Sequence of normal canine COL1A1 CDNA".  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
(FIBRILLAR FORMING COLLAGEN).  
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- SIMILARITY: CONTAINS 1 WVFC DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF153062; AAD34619.1; -  
DR InterPro; IPR000885; -  
DR InterPro; IPR001007; -  
DR Pfam; PF01410; COLFI; 1.  
DR PROSITE; PS01208; WVFC; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.  
FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.  
FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.  
FT DOMAIN 34 92 WVFC.  
FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).  
FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.  
FT DOMAIN 1189 1214 NONHELICAL REGION (C-TERMINAL).  
FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).

FT CARBOHYD 1361 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;  
Query Match 3.0%; Score 391.5; DB 1; Length 1460;  
Best Local Similarity 21.4%; Pred. No. 2.5e-06;  
Matches 379; Conservative 102; Mismatches 645; Indels 643; Gaps 82;  
QY 722 EALHAGNEVPRGEGSGPATVNNSSDTEISIPSPHTAAKDTGONGPKPPA--TLGADGPP- 779  
Db 73 ETKNCPGAQVPPGECPCVCPDGEASPTDOETTCGVEGPKGDTGPRGPRGAGPPRGDIG 132  
QY 780 ---PGPPTPRTSRAPTEPTASEATCAPTP-----PA 811  
Db 133 QGLPLGPPGPP-----GPPGPPGLGNFAFOMSYGYDEKSTGGSVPG 175  
QY 812 PPSAP-----PPVVPKEEKEETAAPPEVEE-----EQPPAAAEELAVDTGKA 858  
Db 176 PMGPSGPRCLPGPPGAPGPGQFQ-----GPPGEPGASGPMGPRGPPGCKNGDDGEA 231  
QY 859 EEPVKSETEAEAGPAKGDAAEAABAGALKAEKKEGGSGRATTAKSSGAPODSDSS 918  
Db 232 GKPGR-----PGERGPPGQAGRLFGTA--GLPGMKHGRFSGLDGAKGDAGP----- 278  
QY 919 ATCSADEVDEAGGDKNLLSP-----RPSLLTPTGDPANASPOKPLDLKQLKQRAAATP 974  
Db 279 -----AGKGEPSGENGAPQCMGPRGLPGERGRGAP-----GPAGARG 319  
QY 975 PIQVTKVHEPPREDAAATKPA--PPAPPPPNLQPSDAPQQPS--SPRKSRSPAPP--- 1029  
Db 320 NDGATGAAGPP---GPTGAPGPPGPAVGAAGKAGGAGPQAGSEGPGQVRCGPPGPPA 375  
QY 1030 -----ADKEAFAAEAKLPGDPPCWTSGLP-FVPPPREVIKASPHADPDPAFSA 1078  
Db 376 GAAGPAGNPGADGQPGKAGANGAP-----IAGAPGFFGARGPSPGQSGPPGKNSG 430  
QY 1079 PPGHPLPLGLDHTARVLPRTPTISNPPPLISSAKHPSVLEROIGAIISOGMSVOLHVPYS 1138  
Db 431 EPGAGNKG--DTGAKGEPGPTGIGPP-----GPAGEGKRG----- 466  
QY 1139 EHAKAPGVFTVNLPLPMDPKLAPFSGVKBQLSPRQOAGPPESLGVPTAQAASVLRGT 1198  
Db 467 --ARPEGPT--GLP-----GPPGERGGPSRGPFGADGAVGPKGP 503  
QY 1199 A--LGSVPGGSITKIP--STRVPSDAITYRGSITHTPADVLYKGTITRIIGEDSPSL 1255  
Db 504 AGERGS-PGPAGPKGSPEGAGRGEAGLPGAKGLT--GSPGSGPDGK-----TGPPGPAQ 557  
QY 1256 D-----RGREDSL-----PKGHVIEGKKGHVLVSIEGGM-----SVTQCSKEDGR 1295  
Db 558 DGRPGPPGPPGARGQAGVGMFGPKGAAGEPKAG-----ERGVGPPGCAVGPAGKDGEA 612  
QY 1296 SSSGPPHETAPKRYTDMMEGRVGRAISSASTIEGLMGRAIPP-ERHSPHLLKEQHHRGS 1354  
Db 613 GAQGP-PGAPAGE-----RGEQGA-GSPGQGLPFGAPGPEAGKPG----- 656  
QY 1355 ITQGIPTSRVYEAQEDYLRREAKLLAREGTPPPPPPSRDLTEAYKQALGLPLKLK--PAHE 1412  
Db 657 --QGVGDLGAPGSGARGERGFPGERGVQGPFGA-----GPRGANGAPND 702  
QY 1413 GLVATVKEAGRSIHPIPREELRHTPELPLAPRLKEGSIQTGP-----LKYDTGASTTGS 1468  
Db 703 GAKG---DAG-----AFGAP-----GS--QGAPGLQMPGERGAAGLPG 736  
QY 1469 KKH-----VRSIGSPR-----TFP--PVHPLDVMDARALERACEYESLKSRRPTA 1515  
Db 737 PKGDRGDAGPGADSGPKGDVGRGLTGTIGPPGAPAGPD-----KGEAGPS 783  
QY 1516 SSSGSIARGAPVIVPELIGKPROSPLTYEDHGAFFAGHLPRGSPVVTMREPTRLREGSL 1575  
Db 784 GPAGTGTARGAPGDRGEGPP--GPAGFA--GPPGAD---GQPAKGEFGDAGAKGD--- 833  
QY 1576 SSKASQDRKL7STPREIAKSPHSTVPEHHHPHI-----SPYEHLRGVSGVDLYRSHIPLA 1631



```
Db 834 -----AGPPGAGTGGPPGIGNVGAPGPKGARGAGP----- 866
QY 1632 FDTSTPRGIPLDAAAYYLRHLAPNPTYPLHYRYPYLDTAALENRQTIINDYIT 1691
Db 867 -----PGATGPGAGRVGP-----PGPS-GNAGPP-----GPPGAGKEGGKARGE--- 908
QY 1692 SOQMHHNTATAMAQADML-----RGLSPRESSLALNYAAGPRGIIDLQVPHL 1740
Db 909 -----TGAPRGVEGPPGPGPAGEKSGPGADGPAGAPGTPGPGIAGQRGVVGL 959
QY 1741 PVLVPT--PGTPATAMDRLAYLPTAPQPFSSRHSSSLSPGPGPHLTXTPTTSSSER 1798
Db 960 PQORGERGPGLPFSGE-----PCKQGPSGTGERGPGPMGPPGLAGPPGESG----- 1009
QY 1799 DRDRERDREREKSLTITTTTVEHAP IWRPGTEOSSGSSG-----GGGSSSRPAS 1853
Db 1010 -----REGS-----PGAGSPGRDGGPGKGDGETGPAGPPGA 1043
QY 1854 HSHAHQHSPISPRTODALQORPSVLHNTGMKGIITAVEPSKPTVLKSTSTSSPVPAATF 1913
Db 1044 PGAPGAPGVGP-----AGKNG-----DRGETGPAGPAG-- 1072
QY 1914 PPATHCPLGGTLDGVYPTLMEVLLPKEAPRVARPERPRADTGH-----A 1958
Db 1073 -----PIG-----PVGARGPAGPQGRGDKGETGEQGDRIKIGHRGS 1110
QY 1959 FLAKPPARSGLEPASSPSKSGEPRLVPVPSGHATARTPAKNLAPHASDPDPAPPASA 2018
Db 1111 GQGGPPGPGSGEGPGSGASGP-----AGPRGPPGSA 1143
QY 2019 SDPHREKTSQKPFISQLELRSLGYHGSYSPEGVEPVSPVSPSLTHDKGLPKHLEELD 2078
Db 1144 GSPGRDGLNGLPGPI-----GPPGPRGTGDAGVGPVPPG----- 1180
QY 2079 KSHLEGELRPKQPGVYKLGGEAAHLPHLPLPESQPSPLLOTAPGVKGHQ----- 2130
Db 1181 -----PPGPPGPPSGGDFSLP-----Op-----PQEKAHDDGGYRYRAD 1215
QY 2131 -----RVVTLAQHISEVITODYTRHHHPQQLSAPLPAPLYSFPGASCPEVLDL 2176
Db 1216 DANVVRDRDLEVDTTLKLSQOIEIRSPESGRKNPAR-----TC--RDL 1258
QY 2177 RRPSPDLYLPPPDHGAPARGSPHSGKRSPEPNKTSVLGG-----GEDGIEPVSP 2227
Db 1259 KACHSDW-----KSGEYWDPNQCCNLDAIKVFCNMETGETCVYPTOP 1301
QY 2228 PEGM-----TEPGHSRSAYVPLLYRDGQTEPSRMGSKSPGNTSOPPAFFSKLTESNSA 2281
Db 1302 QVAQKNWYISKNPKEKRHWYGESMTDGFQFEYG-----GGSDPADVAIQLT----- 1349
QY 2282 MVKSKKQEIKNKLTNHRNEPEYNISQPG 2310
Db 1350 FLRLMSTEASQNIYHCKNSVAYMDQQTG 1378
```

Search completed: September 8, 2001, 14:39:31  
Job time: 433 sec































Qy	734	GECSGPATVNNSSDTSIESPHT-----BAKDTGQNGKPPATLGDG-----PPPG	781
Db	1540	MOOKEKEKQKPEAKEQEPETHPTPEPAETKEPEKAPVAGLPAVTITVVTPEA	1599
Qy	782	PPTP-----PRTRSAPTEPTPASEATGATPPPA---PPSPAPPPVVPKEKE	828
Db	1600	SSAPEKAEEAAEAPSPAGKEPAEPVSEETKLVSEPVSVPEQROSDVPPGEDSRDSQ	1659
Qy	829	EETAAP-----PVEGEQKPPAAEELA-----V	853
Db	1660	DSAAALAPSAPQESAATDAVPCVNAEPLTGTTVSQVSSVDPKPSQPLSKLTQSEEA	1719
Qy	854	DTGKAEPYKSECTB-----BAEGPKAGKDAEAAEATAGALKAKSKKGGSGRAT	904
Db	1720	EEKVKRKPDPTSTEPDATQNGVASEAQPPASEDVANPPVA---AKDRTKNKSRRK	1775
Qy	905	TAKSSGAPODSSSAPTCSADEVD-----EAE-----GG	932
Db	1776	TSVQAAAAASVVEKPVTKSERIDREKLKSSRCEAQKLELKWEAEKITRTASKSSGG	1835
Qy	933	DKNRLSPRPSLLTPGDDPRANASQKPLDLKQKRAAAIPIQVTKVHEPPREDAAPT	992
Db	1836	DTEH---PEPSL---PLSRSRNRNRSVATMTDHSRSPAKEPEVQPRVTKRLERLOE	1890
Qy	993	KPAPP-----PPQNLOPESDAP---QQGSSPRGKRSRSPAPDAKEFAA-----	1037
Db	1891	AVVPPTTPRGRPKPTRRAEEDGHEKKEPAETPRAEGRWSRPSQKSAAGPOGKR	1950
Qy	1038	-----EAKLPGDPCWTSGLFPFVPPPREVI-KASHPADPSPAFSVYAPCHPLPLGLHD	1090
Db	1951	RNEQKYEMAAEAGAQAQSTREGNPKSRGEREASPEKRRDRDPSTDKGPTFPV-----	2004
Qy	1091	TARVVLPRPTTSNPPPLISSAKHPSVLEROIAGISQMSVOLHPVYSEHAKAPVGPVTM	1150
Db	2005	---EVLERKP---PBKTYKSKRGRASTR-----SCMDRAAHORSLEMAAAGAA-2050	
Qy	1151	GLPLPMDPKKLAPFSGVQEOQLSP-RGOAGPESILGYPT-----AQASVLRGTA----	1199
Db	2051	-----DKEAGFAAASPPESSEPOKSGSGSPOLANPADPDRAEESASASTAPEGT	2103
Qy	1200	-----LGSVPGSGITIGIPSTVPDSDAITYRGSLTHG-TPA---DVLYK	1240
Db	2104	QLARQIEQAVQNIATKLEPSSAAAASKGTATATATAASEEPAPEHGHKFAHQASETELA	2163
Qy	1241	GTTRIIGEDSPSRDLRGREDSLPKHGYIYGGKGHVLVSVEGMSVTCSEKDGSRSSGP	1300
Db	2164	AAIGSIISDASGEPENFSAFSPVPGSOTHPRE-----GMEPLGHEAESGILLETGT	2214
Qy	1301	PHETAAPKRTYDMMEGRVGRATSSASISGLMRAIPPERHSPHLKQKHIRGSITQGP	1360
Db	2215	ATESAPQ-----VSALD-----PPGSA-----DTKTRGNSGDV-2246	
Qy	1361	RSYVEAQEDYLREAEKLRKREGTPPPPPSRD-----LTEAYKTQALGP	1404
Db	2247	-----QEAGSKAEVTPPRKDKGRQKTRRRKNANKKVVAITETRASEAQT	2294
Qy	1405	LKLKPAHEGLVATVEAGRSIHEIPRELHTP-----ELDAPRPLKEGSIQTGTP	1456
Db	2295	QSESAAEEATAATPEAQO--EEKPSRPPSPAECTFDPKSTPPASLSQENSAAEKTP	2352
Qy	1457	LKYDTGASTTGSKKHVRSIGSPGRTTPPVHPLD--VMADARALERACYEESLKRSPGT	1514
Db	2353	CK-----APVLPALPPLSQPALMDGQPAQKFKVHHSIESDPVT	2390
Qy	1515	ASSGG-----SIARGAPVIVPELKGPRQP-----LTYEDHGAPPAFAGHLPRGS	1558
Db	2391	PPSDSGIPPTTPTLVITIAKLPPPVIPG-GVPHQSPPPKPVTEWITRQE-----E	2437
Qy	1559	PVTRPPTPRLQSGSLSS---SKASQDKRLTSTPREIAKS-----PHST--V	1600
Db	2438	PRAGQSTPSALPPDPTKASDMDTSSSTLRKLMDPKYVSATCVTSTSVTTTAAIBPVPACPL	2497

RESULT	5
076891	
ID	076891
AC	076891

Qy	1601	PEHHPHIPISPYEHLRGVS-----GVDLYRSHIPIAFDPTSIPIRGIDPLDAADAAAYLPRHLL	1655
Dy	2498	QEAPAPPDCDFKHPLEBVGAAAAPVNAQTQASEVPVAADKEKV---APV-----I	2543
Qy	1656	APNPTYPHYLPYIRGYPDTAALENROFIINDYITTSQOMHNTATAMAQRADMRLGLSP	1715
Dy	2544	APKIT-----SVSRMPVSDLENLSQKI-----TLAKPAQOTLTGL--	2579
Qy	1716	RESSLALNYAAGPRGIIDLISOQPHFLPVLPPTPGTAMDRLAYLPTAPOFSSRRHSS	1775
Dy	2580	-----VSALTGLVNYSLV-P-VNALKGPKVGSVATLKGLVS-----	2613
Qy	1776	PLSPGGPTHLT-K-PTTITSSEBERDRDRDREREKSILTSIT-----TYEHAPIWRP	1829
Dy	2614	--TPAGPVNLKKGPVNWLTG-----PNVLIUPTVSATGVFNAA-P-GP	2653
Qy	1830	-----GTEOSSSGSGSGGGSSRPASHAHQHSPISPRTODALQQRPSSLHNTGMK	1884
Dy	2654	VTAACGVTTATTGAATGAVTAPAAGKQORASSNENSFRHFPGSMSVIDDRPA---DTG-S	2709
Qy	1885	GIIATAVEPSKPTVLRTS---TSSPVRPAA---TPPPATHCPGLGTLDGVYPTLMPEVLL	1938
Dy	2710	GAGLRVNTSEGCVLLSYSGKTGEGPORISAKISQIAPPASAMI-----	2752
Qy	1939	PKEAPRVARPERPADTGHAFLAKPARSGLEPASSPSKSEPRLPYPVYSGHATIARTP	1998
Dy	2753	--EFQOSVKSQWKADS-----IPTQSAPKGPQTFSAFANAHAHS TLVLT -	2796
Qy	1999	AKNLAPHASDPDPAPPASADSPHREKTQSKPFSTOELESLRGVHGSSYSPEGVEVPSP	2058
Dy	2797	-----AQYN-----ASP	2804
Qy	2059	VSSPSTHTDKGLPKHEELDCKSHLEGELRKQPQGVKLGGEAAHLPHLRPLPESQSPSSP	2118
Dy	2805	VIS-SVKTRD---PSLEKPEPETHLSYSTPVTGGTVKVLTOGITNTPPV-LVHNQLVLTP	2858
Qy	2119	LLQTAFCVKGHQHVVTLAOHISEVLT-----QDYTRHHPOQLSAPLPAPLY	2164
Dy	2859	-----SIVTNKKLADPVTLKIETKVLOPANLWGLPTLTHPHHPALPSKLPAEVN	2906
Qy	2165	SFP-GASCP-----VLDLRRPPSDLYLPDPDHGAPA---RGSPHSEGGKRSPENKTSVLG	2216
Dy	2907	HVFSGSPTPADRTIAHLATKPDTHSPREPTGPTGLFFRPCHPSSITTSTALSTNATVMLA	2966
Qy	2217	GG---EDGIEPVSPPGMETPGHS-----RS	2239
Dy	2967	AGIPVPOFISSIHPEQSVTMPHSHISTQT VSLGHL SQGEVRMSTPLPSITYSIRPETLHS	3026
Qy	2240	AVYPLLRYDGEQETEPGRMSKSPGNTSOPPAFFSKLTESN-----SAMVXSKQKE	2289
Dy	3027	PRAPFOQIEARAPQVGTGPQATGTPVALATQHPPPEEVHYHLPVARAANPVOSEVLV	3086
Qy	2290	INKKLATHNRNEPEYINISOFGETEINFMPAITGLMTYRSOAVEHASTNNGLEAIIRKA	2349
Dy	3087	MOSEYRLHPVTVPR-----DVRIMVHPHTAVSQPRATEGVGVKVPANKAPQOLVKEA	3140
Qy	2350	IMGKYDOWEEESPPLSANANPNI NASASUL PAAMPITAADGRSDHILTSPGGGGGKAKV-SGR	2408
Dy	3141	-----VKTSDAKAVPAPVPVPVPVPT-----PAPPHGEARILTVT	3178
Qy	2409	PSRKA K SAPGLASGRDP-----SVSSVHSEGCNRRRTPLTN-RVYEDRPSSAGSTPF	2462
Dy	3179	PSOLOGLPL-----TPVVVTHGVQIVHSSGELFOERYKGDVRYTHAPAQOULTHQF	3231
Qy	2463	PY-NPLMRLOQVMASPPPPGLPAGSGLAGP	2494
Dy	3232	PVASSISLASRTKTSAOVPECEPLOSOSAOP	3264

PRELIMINARY; PRT; 5327 AA.











QY 774 GADGPPGPTTPTTPTSRAP--IEPTPAEATGAPTTPAPPSPAPPVVPKKEEBT 831  
Db 1329 DAK-----PPTGASFSQAESNVDPEDS-----TQPLSKPAQKSEANEPEAKPDAT 1377  
QY 832 AAAPP-----VEGEQKPPAAEELAVD---TGKAEPPVKSECTEEAEFGPAKGDAA 883  
Db 1378 ADAEPDANKAEAPESQPPASEDELEVPPVAAKDKKPNKSKRKTTPVQAAAVSIVEKPV 1437  
QY 884 EATAEGALKAEEKGSGRATTAKSSGAPQSDS--SATCSADEVDEAEQDKNRLLSPR 942  
Db 1438 TRKSEKIDREKLKRSNPRGEAQKLELKMAEAKITRTASKNSAADLE-----HPEP 1489  
QY 943 SL-LTPT-----GDPRANAPOK-PLD-----LKOLKORAAAIPPIQVTK 980  
Db 1490 SLPLSTRNRNRSVYATMGD--HENRSPKVEPVEQPRVTRKRELERELQEAAYV--TTPR 1546  
QY 981 VHEPR-----EDAAPTKPAAPP-----PQNL-----QPESDAP 1012  
Db 1547 RGRPPKTRRRDEEENEAKPAETLKPEGWRSRQKTAAGGGPQGGKKNPKVDAT 1606  
QY 1013 QPGSSPRG-----KSRPAPPADKEAFAAEAQKLPGBDPPCHTSLGPPVPPREVIKASP 1067  
Db 1607 REATEVQPIQVKESMEPKAAEEAGSEQDRKD--AGTDKNPPTETAPEVVEKPK 1664  
QY 1068 HAPDPSAFSAPPGHPLPLGLHDTARPVLPRTTISNPPPLISSAKHPSVLERQIGAISQ 1127  
Db 1665 -APERNSKS-----KRGSRNSRLAVDKSASLKN---VDAVSPRGAQAQAGERES 1711  
QY 1128 GMSVOLHPVYSEHAKAPGVPMGLPLMPD-----KKLAPFSG-----VKOEQL-- 1172  
Db 1712 GV-VAVSPKESPOKEDGLSSQLKSDPVDPDKEPEKEDVSASGSPREATOLAKQMELEQ 1770  
QY 1173 -----SPRGAQGPESLGVTAQEAASVLRGTALGSV----- 1203  
Db 1771 AVEHIAKLAESASAYKADAPLEGIA--PEDRDKFAHQASETELAAGIISINDISGPEE 1828  
QY 1204 -----PGGSITKIGIPSTRV-----PDSAITYRGSIHTGTPADVLVKGTI--TRIIGED 1250  
Db 1829 NFPAAPPYGESQTDLQPPAGAQAQPSE-----EGMETDEAVSGILETEAAYTES 1878  
QY 1251 S-----PSRLORGREDSPKGVHVIYEGKKGHVLSYEGGMSVTCQSKEDGRSS 1298  
Db 1879 SRPPVNAPOPSAGPTDTKEARNSSETSHSVPDEAGSK-----EVEVTLVRKDKGRQ-- 1930  
QY 1299 GPPHETAPKRIYDMWEGRVGRAISSASIEGLMGRAIIPERHSPHHLKEQHIRG---SI 1355  
Db 1931 -----KTRRRKRNNTNKKV-----VAPVESHVP-----ESNOAQGESPA 1966  
QY 1356 TQIGIPRSYVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYKTQALGLPLKPAHEGLV 1415  
Db 1967 NEGTTVQHPEAPQ-----EEKQSEPHSTPPQOSCTSDLSKIPSTE----- 2006  
QY 1416 ATVKBAKRSIHPIREELRHT-----PELPLAPRLKEGSIQTGGTPLYKYDTGASTGSKK 1470  
Db 2007 -----NSSQBISEVERTPTKASVPDPLPPPQP-----APVDEEPOA-----R 2044  
QY 1471 HDVRSILGSPGTFP--PVHPLDVNADARALACRYESLSKRPCTASSSGGSIARGAPV 1528  
Db 2045 FRVHSIIESDPVTPPSDPSIPIPTLPSV-----TAAKLSPPVAVSG--- 2084  
QY 1529 IYPELGKPRQSP-----LTYEDHGAPFAGHLPRGSPVTMREPTPRLOEGSLSS---SK 1578  
Db 2085 -----GIPHQSPPTKVTEWITROE-----EPRAQSTSPALPPDPTKASDVDT 2127  
QY 1579 ASQDKLSTPREIAKS-----PHSTVPEHHPIPISEYHLLRGVSGVDLYR 1625  
Db 2128 SSTLRKILMDPKYVSATSVTSTVTTAIEAPVSAAPCLHEAPPVPPVD-----SKKPLEE 2181  
QY 1626 SHIPLAFDPTSIPIGIPLDAAAAYLPHRLAPNPPYPLHYPLIRGYPDYTAALENRQTI 1685  
Db 2182 KTAPPVTNNSEIOASEVLVAADKEKVAVPVIAPKIT-----SVISRMPVSDILENSOKI 2234

QY 1686 INDYITSOQMHHNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVP 1745  
Db 2235 -----TLAKPAQTITGL-----VSALTGLVNVSLVP--VNALKG 2267  
QY 1746 PTPGPTATAMDRLAYLPTAPOQFSSRHSSSPSLSPGGPHLTKPTTTTSSSERERDRDRD 1805  
Db 2268 PVKGSVTTLKLSVS-----TPAGPVNVLKGPV----- 2294  
QY 1806 RDREREKSLISTTIVEHAPI-----WRPGTQSSSGS-----SGSS 1841  
Db 2295 -----NVLTGPNVLTTPVNAATVGTVNAAGFTVNAASAVNATASAVTVTAGAVTAAS 2347  
QY 1842 GGGGGS-----SSRPASHSHAHOSPTSPRTQDALQORPSPVHLNHTMGKIIT 1888  
Db 2348 GGVATTGTVTMAGAVIAPSTKCKORASANENSRPHGSMPIVDDRPA---DAG--SGAGL 2403  
QY 1889 AVEPSKPTVLRSTS---TSSPVRRPAA---TFPPATHCPLGGTLDGVYPTLMEPVLPLPKEA 1942  
Db 2404 RVNTSEGVLLSYSGQKTEGPORISAKISQIPPAS-----AMDTFQGSVSKSQVKPDS 2457  
QY 1943 PRVAP--ERPADTCHAFLA-----KPPARSGLPEPASSPSKSGSEPRPL--- 1984  
Db 2458 VTASQPPSKGPOAPAGYANVATHSTLVTAQYNASPVISVVK--ADRPSL--EKPEPIHLS 2515  
QY 1985 -----VPPVSGHATITRTPA----- 1999  
Db 2516 VSTPVTQGTGKVLVLTQINTPPVLVHNLVLTPTSIVTTNKKLADPVLTKIETKVLPANL 2575  
QY 2000 -KNLAPHIASDPPAPPASASDP--HREKTOSKFP--SIQELELSRLGYHGSYSPEGVE 2054  
Db 2576 GSTLTTPHH---PPALPSKLPTEVNHVPSPGSPADRTVSHLAAAKLDAHSRPSGPGPS 2631  
QY 2055 PVSVPSSSLTHDKGL-----PKH--LEELDOKSHL- 2082  
Db 2632 SPFRASHPSSTASTALSTNATVNLAAIGIPVPOF ISSIHPEQSVIMPPHSITQTVSLSHLS 2691  
QY 2083 EGELRPKPG-----PVKLGGEAAHLPLRPLPESQSSPPLLQATAPGVKGHORVVTLAQH 2138  
Db 2692 QGEVRMNTPLPSIYTSIRPEALHSR-APL---QP-----QOIEVRA-- 2730  
QY 2139 ISEVITQDTRHHPOOLSAPLPAPLYSPFGASCVPDLRRPPSD---LYLPPPDHGAPAR 2195  
Db 2731 -----PORASTPOAP-----AGVPALASQHPPEEVHVLPLVARATAPVQ 2771  
QY 2196 GS-----PH-----SEGKRSPEPNKTSVLGGEDG 2221  
Db 2772 SEVLVMQSEYRLHPTVTPROVRKIMVPHVTAVSEQPRADGVVVPKVPASKAP---QOPG 2827  
QY 2222 IEPVSPPEGEMTEPGHSRAV-----YPLLYRDCQEOTEPSRMGSKSPCNTSQ-----PP 2269  
Db 2828 KEAAKTPDAAKAAPTTPAPVPVPLPAPAPAPHGE---ARILVTPSNQLOGLPLTTP 2883  
QY 2270 AFFSKLTSNSAMVSKKQKQKINKNLNTNRNE--PEYNISQPGTEIFNMPAITGTGLMTY 2327  
Db 2884 -----VVVTHGVQI-----VHSSGELFQERYGD-----IRTY 2911  
QY 2328 RSQAVQEH---ASTNMGLEALIRKALMKGDOWEE--SPPLSANAFNPLNASASILPAAMP 2382  
Db 2912 HPPAOLTHQTQPPAASVGLSPRTKTAAGQPPPEGEPLPPQVQSTQPAQAPPCPPSQ-- 2970  
QY 2383 ITAADGRSDHTITSPGGGKAK---VSGRPSRRAKSPAPGLASG--DRPP-----SVSSV 2433  
Db 2971 -----LGQGGQPPSSKMPQVSOEAKGTQGVQBPRLPAGPANRPPPEPHVQVORA 3019  
QY 2434 HSE-GDCNRRTPLNRVWEDRPSSAGSTFPFYNPLIMKRLQAGVMASSPPGLPAGSGPLA 2492  
Db 3020 QAETGPTSFPSVPSVSMKPDPLPSLPTOTAPKQPLFVPTTSG-----PSTPPG---LV 3069  
QY 2493 GPHHAWDEEPK---PLLCQ 2509  
Db 3070 LPHTFQBPAPKQDSSPHLTSQ 3090







QY 1847 ---SSRRPASHSHAH-----QHSPISTQDALQORPSVLHNTGKMGITAV 1890  
DB 1342 MPNSSAAAAAARAAAPASSVSLRSQSLHPVQPSLSHPS---SSALSAAAAAV 1398  
QY 1891 -EPSKPTVLRLSTSSPVRAATFPATHCPL-GGTLGGVYP-----TLMEPVLLP 1939  
DB 1399 AERDRHALMRQOS-----PHMTTPPVSNASLSPSKMYAPQPGQGLGTSPPPHLRP 1452  
QY 1940 KEAPRVAR-PERPRADTGHAFKAPARSGLPEPASPSKSPRPLVPVPSGHATTARP 1998  
DB 1453 GASPPVIRHPQML-----PLLIAPGGGIPQIGVHP 1484  
QY 1999 AKNLAPHASPOPPAPPASADPHREKTSQKPSIQLELRLSLGYHGSYSPEGVPSVP 2058  
DB 1485 GQSPYH-----PLHPSVFSYSPHH-----PPN-----SPGYAPYGP 1518  
QY 2059 VSSPSLTHDKGLPKHLEELKSHLEGELRPQGPVKLG-FAAHLPHLRPL-ES 2112  
DB 1519 -----GFPAYMKP-----PPQGLDPAAVMAAHAGLQGPQOMRQD 1557  
QY 2113 QPSSSPLLQATPGVKHORVTLAQHISEVITQDY-----TRHHPQQLSAPLPAP 2162  
DB 1558 EQNAAAAAQAARAKOHAQAAAAAQAQKAPQOQPGMPPNKPPTPKTPQCGGMP-- 1615  
QY 2163 LYSFGASCPLDLRRPPSDL-----YLPDPDHG---APARGSPHS-----EGGKR 2205  
DB 1616 -----PCMGGPGTGLPGGAYPCSHMPGYPOGPHGSPAPQDQPHGLKPSHMDALRA 1671  
QY 2206 SPEPNKTSVLGGEDGIEPV-----SP-----TEPGHSRSA 2240  
DB 1672 HAHSAANSAGMGHGHTEPLIDIEPDPEIPSPHNPGRGSPPEAKPDDTECHRSQSA 1731  
QY 2241 VYPLLRYDEQTEPRMGSKSPGNTSQPAFTSKLTESAMVSKKQKINLKNTHNRN 2300  
DB 1732 IFVRHIDRGDYNSTR-----TDLIFKPVADSKLA---RKREDRKLAEKERE 1777  
QY 2301 EPEYNISQPGTEIFNMPAITGTGLTYRQAOQEHASTNMGLEAIRKALMGKYDQWES 2360  
DB 1778 RROOQOQO-----QOQOQOQOAAAAQAQAKKA-----ELK 1811  
QY 2361 PPLSAN-----AFNPL-----NASASL-----PAAMPI 2383  
DB 1812 PPVADPPALRQLSEYARPHVAFRELEIKNAQAAAAAQSRLDPHMYRRGTHPSQFPL 1871  
QY 2384 TAADGRSDHTLSPGGGKRAKYSGRSSRKAKSPAGLASGDRPPSV 2430  
DB 1872 YANPAISQMERERLGIPPPHVGLDPEGHVWRMPPEAGFOLPPNV 1918

RESULT 8  
Q9P3J0 PRELIMINARY; PRT; 2649 AA.  
AC Q9P3J0  
DT 01-JUL-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CONSERVED HYPOTHETICAL PROTEIN.  
GN B7F21.40.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL369501; CAB97476.1; -;  
DR InterPro; IPR001005; -;

DR PROSITE; PS50090; MYB 3; 1.  
SQ SEQUENCE 2649 AA; 294398 MW; 36FB50BF3917F2AB CRC64;  
Query Match 4.2%; Score 552; DB 3; Length 2649;  
Best Local Similarity 19.3%; Pred. No. 1.7e-19;  
Matches 406; Conservative 268; Mismatches 771; Indels 658; Gaps 92;  
QY 60 QQORRRSLLSEFPQG--NERSOELHLRSHESLYLDELKSEM--EFIESKRPRLELLPD 115  
DB 848 KPVQKPAQKLDLEPPRVSESEVAPKPEPKVPSEPKKAPPELLPEKLPPEQIPR 907  
QY 116 PLLRPS-LLATGQAPGSEDLTKDRSLTGKLEPVSPSPPTDPE 159  
DB 908 POPQPOQPOPOQAQAQAPPOLLTAEAPKSVTEPEREAREAEVQEAETQEQAP-TERQ 966  
QY 160 LELVP-----PRLSKEEL-----IONMDRVDREITTWYEQOISK 193  
DB 967 LLAIPVTEENEETSFLRPEPQPKMEEMDTGGSLPLPTVVEHKVVDQDVMVD----- 1020  
QY 194 KKKQOOLEE---AAKPEPEKPVSPPIESKHSRLSVQIYYDENRKAARAH----- 242  
DB 1021 -----VDDEGRDALQP--PCKPVLRLDVSQGESVSQFPRFNQPTSDGASSRMD 1072  
QY 243 -----RILEGLGQVPELPLYNQPSDTRQYHENIKINOAMRKK--LILY 283  
DB 1073 EERTDDASVSGSVHRELSTTPPTDELFPN-----VRPWHOSSKVRKLSSESPGFGAF 1128  
QY 284 F-----KRNHARKQWKFCORYDQLEMEALEKKVERIENNRPRRAKESKVR 331  
DB 1129 FMGNIRDMAETAKEQHEA---KQVRYTNDAYLRF-----TLSDDP--TAVKSR--- 1173  
QY 332 YIEKQPPETRKQRELQERMOS---RVQRGSGLSMAARSEHVSII--DGLSE-QENLE 386  
DB 1174 ---NQSQDKDKVSGKGSGHSGNDHKGEGRRTTSRFTSLDVEYAIQESIREAQEKKE 1230  
QY 387 KOMR-----OLAVIPPMY--DADQRIKFINNGLMADPMKYVKDQVYNNM- 432  
DB 1231 REERAQREKRYTDEKAVIFEMITWDEKHQLFVDSGLL--PLE-----KLIVETQWAVP 1283  
QY 433 -----SEQKETFREKPMQHPKFNGLIASFLERKTVAECLVLYLTKKNKYKSLVRSY 487  
DB 1284 YHVNFTAEAEKFEKAYLEYPKQWKGIAHELPNRDPHSVIQIYVYAKKRELNLKRLKQ 1343  
QY 488 RRRGKSQOQOQ-----QOQOQOQOQOQPMPPSSDEKDEKEKEKEKEEKEPEV 538  
DB 1344 RRRKKRGKQKYNALVSELGNPENETTDNQENGENSERRQOQPRRAAAPSNGHEATPNA 1403  
QY 539 END-----KEDLLKKTDDTSGEDNDEKEAVASKGRKTANSQGRKGR- 582  
DB 1404 DSDGATPATPGRRRAGTTTEPKNDSGAE---KPEGVKKGRRARQPKADKEKVPKPAQ 1460  
QY 583 -----TRSMANEANSEEAITPQOSAE----- 603  
DB 1461 AIAPTTPAASGKGTNARSNSRTRVQNPEWMPKTPAELGARGVQNDPMFDVPPGSMOPPLA 1520  
QY 604 -----LASMELNE----- 611  
DB 1521 PAQQPTLASEPAPPTLASTTISEVNAAPSLRPEPPAPPASLPTTFEIGOSSGPERTRTQ 1580  
QY 612 ---SSRWTEEMETAKKGLLEHGRNNSAIARMVSGKTSVQCKNFYFNK--RQNLDELIIQ 667  
DB 1581 QASSYWSVSESNDFPSLLRSFGTDSAIANHMGTKTVQVNVRYLQRKKEGQEWQIAT 1640  
QY 668 QHKLMEKERNARRKKKAPAAASEEAPPPVVEDEMEASGVSGNEEWEVEAEALHAS 727  
DB 1641 EADLK--KQGERRRPPPTPSAGPR-----KRYDVPSSSGHRPLAAAEAE- 1684  
QY 728 GNEVPRGEGSGPATVNS-----SDTESIPSPHTAAKD--TGQNGPKPAPTLGAD-GPPPG 781  
DB 1685 -PQPAKSEAAPANQPFSSRFQVTPIQASVLTLLVQSRSTVMPAPLPAPVPAASASAA 1743  
QY 782 PTPPRTSRAPTEPTASEATCAPTPPPAPPSPAPPPVPKPEKEEETAAAPVVEEGE 841















QY 1668 -YLIRGYPDTAALNRQTIINDYITISQMHNTATAMAQADMLRGLSPRESSIAL---N 1723  
Db 1211 EKLIPGLSPVARGSGNSVGMEDGPRRRRGRQAQDK-----PPFRRLQAOREN 1266  
QY 1724 YAAGPRGIDLSQVPHLPVLVPTTGTATAMDRLAYLPTAPQPPFSRRHSPLSPGGPT 1783  
Db 1267 AARGSEKPSLT---LPASAP---APEEA---LTTVTVPAPRAAAKSPDLSNQSD 1315  
QY 1784 HLTKPTTSS-----SERDRDRE-----RD 1805  
Db 1316 QANEWEATASESDTSERRGDKAPPVLLTPKAVGPGGGGAVPGISAMSRGDLQS 1375  
QY 1806 RDREREKSLTSTTVEHAPIRWPRTEQSSSGSGGGG----- 1846  
Db 1376 RAKDLSKRSFSSQRMERQNRPPGKAGSGSGSGGGGPGGRTGPGRGDKRWS 1435  
QY 1847 ---SSRRPASHAHQSPISPTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSST 1904  
Db 1436 PKNRSRPPPEERPPGLPLPPPPSSSAVRLDQVH-----SNPAGIQQAALQ 1482  
QY 1905 SPVRAATFPATHCPLGLTGLDGVYTLMEPVLLKEAPRVARPERPRADTGHAFLAKPP 1964  
Db 1483 LSSROGSVTAPGGH-----PR-HKPGPPQAPGQPS---PRPP 1515  
QY 1965 AR-----SGLEPASPSKSGSEPRPLVPVSG-----HATIARTPAK----- 2000  
Db 1516 TRYEQRVNSGL---SSDPHFEECPGPMVGVGTGPRDSAGVSFPFKRRERPPKPELLO 1572  
QY 2001 ---NLAPHIAS-----PPPPAPPASAD-----PH-----REKTSKPSISQBLE--- 2037  
Db 1573 EBSLPPHSSGFLGSKPGCGQAESRDTGTEALPTWNLHTATSRKSRPSSMPEWPM 1632  
QY 2038 ---LRSLGYHGSYSPEGV-----PVSPVSSPSLTHDKGL-----PKH--- 2073  
Db 1633 EPLSFEDVAGTEMQSQSGVSDSGVSGPCQSRSP-----DGLGKAAGGPKPRPG 1688  
QY 2074 ---LEEDKSHLEGRLKPGVPLKGEAAHLPHLPLPESQPSPLQATAPVKGKH 2129  
Db 1689 GSSPLNAVCEGPPGSEPPRRPPAPHPDGRKELPREQPLP-----PGPIGT 1735  
QY 2130 QRVVTLAHSIVITQDTRHHQQLSAPLAPLYSFGASCPLVLDLRPPSDL----- 2183  
Db 1736 ER-----TDRCTEGPIRPS-HRPGPPVQF-GTSDKDSLDVLWGDLSLKAKE 1784  
QY 2184 ---YLPDPDHGAPARG-----SPHSEGGKRSPE-----PN 2210  
Db 1785 LTASVTEAIPVSRDWELPLSAASAREPQSKNLDGCHVPEPSSSQRLYPEVFGSAGPS 1844  
QY 2211 KTSVLGGGED-----GIEPVSP-----PEGMTEPGHRSVAV-YPLLYRDG 2249  
Db 1845 SSQISGGAMDSQLHPNSGFRPGTSLHYPYRSOPLYLPPGAPPALLSGVALKGQFLDF 1904  
QY 2250 EGTEPSRSGKSPGNTSOPPAFF-----SKLTESNAMSVMKSKQBINKLNTHRNE 2301  
Db 1905 STMQATELGPAGVLYPPPSFLYSFAPCFSPPLDTSLLVQV----- 1948  
QY 2302 PEYNISQPGTEFNMPAITG---TGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDOWEE 2359  
Db 1949 ---DLPSP-SDFYSTPLPLOGGSGFLPSGAPAQ-----MLLPWDSQLPVVNFSGSLPP 1998  
QY 2360 SPPLSANAFNLNASSALPAAPIITAADGRSDHTLTSPGGGKAKVSGRPSRRKAKSPAP 2419  
Db 1999 APP---PAPPPLSLPVGALOPPSLA-----VRPPAPATRVLPSPA- 2038  
QY 2420 GLASGDRPPSVSSVHSE-----GDCNRRPTLTNRWEDRPSSAGS 2459  
Db 2039 ---RPPFASLGRAELHVPKLPFDQYQKLSSNLGPGGSSRTPTGTSFSGLSRLKA 2092  
QY 2460 TPFYNPILI---MRLOAGVMASPPP-----PGLPAGSGP---LAGPHHAWDEEP 2502  
Db 2093 TFSYSGVFTQRVDLYOQASPPDARLWPKPWERTGLPPREGPSRRAEPEPSRGDKPEP 2151

RESULT 11  
Q9VRV3  
ID Q9VRV3 PRELIMINARY; PRT: 3080 AA.  
AC Q9VRV3;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG10115 PROTEIN.  
GN CG10115.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Beres P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Foster C., Gabrielista C.C., Ferraz C., Fertier S., Fleischmann W.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003561; AAF50647.1;  
DR Flybase: FBgn0035712; CG10115.  
DR InterPro: IPR001487;  
DR IntronPro: IPR001487;  
DR Pfam: PF00439; bromodomain.1.  
DR PRINTS: PR00503; bromodomain.  
DR PROSITE: PS00455; AMP\_BINDING.1.  
DR PROSITE: PS0014; bromodomain.2;  
DR SMART: SM00297; bromo.1.  
SQ SEQUENCE 3080 AA; 338176 MW; 4602054387DE2C12 CRC64;

Query Match 4.0%; Score 533; DB 5; Length 3080;  
Best Local Similarity 18.6%; Pred. No. 1.8e+18;  
Matches 561; Conservative 385; Mismatches 1041; Indels 1022; Gaps 138;















OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=96312450; PubMed=8698236;  
RX Yotov W.V., St-Arnaud R.;  
RT "Differential splicing-in of a proline-rich exon converts alphaNAC  
RT into a muscle-specific transcription factor.";  
RL Genes Dev. 10:1763-1772(1996).  
DR EMBL; U48364; AAB18734.1; -  
DR EMBL; U48363; AAB18732.1; -  
DR MGD; MGI:106095; Naca.  
DR InterPro; IPR002715; -  
DR InterPro; IPR003037; -  
DR Pfam; PF01849; NAC; 1.  
DR Pfam; PF02094; TS-N; 1.  
SQ SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

Query Match 3.9%; Score 520; DB 11; Length 2187;  
Best Local Similarity 19.7%; Pred. No. 5.5e-18;  
Matches 448; Conservative 219; Mismatches 782; Indels 830; Gaps 93;

QY 738 GPATVNNSDTESIIPSTEAA-----KDTGQNGKPPATLGDGPP---PGPPT 784  
DB 3 GEATETVATEQELPOQAETAVLPMSSALKVAAVQCGPPTPPSSILGQQSQPIVTAHQPS 62  
QY 785 P-PRRTSRAPIE---PTPASEA---TGAPTP---PP-----810  
DB 63 PLPSSVSTPEFVPPAQPIATAALPSGTAPPTPFLPHLIGPPISPAALALASPMIGLA 122  
QY 811 -----APPSAPPVVPVKEEKEETAAPVE-----838  
DB 123 QKGARSSAPLSLVALPHSVQKSSVCPPHPLTSPSAGAGALGALTASIPLEPKTSTS 182  
QY 839 -----EGEEKQPP-----AAELAVDTGKAEPVKSE 865  
DB 183 QVPSQGLTNLKGATPCPDVVRAPFSLHENPLASVQPLGMLSCPTLSNTSPVKGVPISSA 242  
QY 866 CTEAEEGPAKGDAAEAETAEGALKAE-----KKEGGSGRAT 904  
DB 243 LTQSLRLNLKGVSPPARNTAAPSILAPSTSLGCHPLLLHSSVDSPIQPGQSGGLAV 302  
QY 905 T-----AKSSGAPQSDSSATCSADEVDEABEGDK- 934  
DB 303 SNPTSVGHSGIAACPPERCVVVPALPSRLAVDSCAAPSDDKGSAVTNELCSPPGSSNV 362  
QY 935 -NRLISPRSLTTPGDRANASPOKPLDLKOLKORAAAIPIQVTKVHEPP-----985  
DB 363 AGTSLSPKASLV-----PKGSNVALQPL-----VTQVPASQKTGLKEIPVSCIGATH 409  
QY 986 ----REDAAPTAPAPPAPPONLQESDAPQPGSSPRGSRSPAPPADKEAFAAEAK 1041  
DB 410 HALDNPSAISVAPATHVPPT-----SSGLVSKDPAASVTSILVPPAAHKQ 455  
QY 1042 LPGDPCWTSGLPF-PVPPREVIKASHAPDPSAFYAPPGLPLGLL-----HDTARPVL 1096  
DB 456 FPAPPASATLGVVSPLEATEGLK---NLPIASLVNMGAPVSPAQAGLPTKRKDTTLQPLA 512  
QY 1097 P-----RPPTISNPPPLISSAKHPSVLEIROIGALSQMSVOL 1133  
DB 513 PIALKESPPSSASLEVLSEDTVTKTGGPAPVVRFA-----IAGVATTSLSRA 563  
QY 1134 HVPYSE-HAKAPVGVPTMGLPL-----PMDPKLAPFSGVKQBLQPRGQAGPPE 1182  
DB 564 DSPPAVIRADSCVSNPTVSQPLKRSVTDPMAPRTAKTAPST---TSPLVPLASEGCPV 620  
QY 1183 SLGVPTAQEASVLRGTALGVSPGGSITKGP-----STRVPSDAITYRGS 1229  
DB .621 ASSMALSPQNASVSETALALSP--EIPKSVFPDPLAEISFNARKVDVASHMESSGS 678  
QY 1230 THGTP-ADVLYKGTITRIIGDSSPRLDRGDS-----LPKGHVI 1269  
DB 679 ROGHPDASVTAKTGVVCL-----ADSSLDTSVSAKGSALSGASSPLYPLEVSFLPEAGLA 734

QY 1270 YEGKKGHVLSYEGGMSVTOCSKEQG--RSSGPPHETAA-----1306  
DB 735 VOGPKGSL---NKLSPTPSSKGAPVSTGAPSPKGAIVPTSESSISSKQVPAEILPS 790  
QY 1307 PKRTYDMMEGRVGRGAISASIEGLM-----GRAIPPERHSPHLKEQHHI---1351  
DB 791 POKTEPVTASRLISAVQSPKVDPIINSDDVTPTSPPKTSATAVPKDTSATLSLKSVPAVTSL 850  
QY 1352 -----RGSITQGIPT-----1361  
DB 851 SPPKAPVAPSNEATIVPTETISLKNALAAATPKETLATSIPKVTSPSPQKTPKSVSLKG 910  
QY 1362 -----SYVEAQED---YLREKALLKE-----GTPPP--pP 1388  
DB 911 APAMTSKKATEIAASKDVSPSQPFKEVPLQLQHVPTSPKPSVSDTLGALTSPPKKPP 970  
QY 1389 PSRDLEAYKTOALGLKLIK-----PAHEGLVATVKEAGRSIHEIPR---EELRHTPEL 1439  
DB 971 ATLAETPTYPKSKPKPAASKKTPATPSPPEGVTAVPLEIPPCSKKAPKTAAPKESATSSS 1030  
QY 1440 PLAPPLKEGSITQ-----GTPLKYDTGASTTGSKKHDSRLSLIGSPQRT-----1483  
DB 1031 KRAPKTAVSKETPSKGVTAVPLEISLPLKETSKSATPGEKS-----ASSPKRSKPTAGP 1084  
QY 1484 -FPP-----VHPLDVMADARALERACEESL-----KSRPGTA---SSSGGSJAR---1524  
DB 1085 KETPPGGVTAVPPELISLPKETPQNATPNESLAASSQKRSKTSVPKETPPGCVGTAMPLE 1144  
QY 1525 --GAPVIIVPELKGKPROSPLTYEDHGAPFAGHLPRGSPVPTMREPTPRLQEGSLSSSKASQD 1582  
DB 1145 IPSAPQKAPKTAVPKQIP-TPED-----AVTILAGSPLSPKK-----ASKTAAP 1187  
QY 1583 RKLSTPR-----ETAKSPHST-----VPEHHPHIPISPYEHLLRGVSGV 1621  
DB 1188 KEAPATPSVGVIASGEISPSPKKTSKTAAPKENSATLPPKRSKPTAAPKETPATSEGV 1247  
QY 1622 DLYRSHIPLADPTSIPIRGIPLDAAAVYLRHLAPNTYTPHLYPPYLRIGVDPDFAALEN 1681  
DB 1248 TAVPSEISPS-PPTPASKGVVP-TLTPKAPNALAESPASP-----KKVPKTAAPEE 1297  
QY 1682 RQTIINDYITSQOMHHNTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVHLP 1741  
DB 1298 TST-----TPSQKIPKVG--PKRAS-----1317  
QY 1742 VLVPTPTGTPATMDRLAYLPT-----APQPFSSRHSSSPSLSPGPGTHLTKPTTT 1791  
DB 1318 -ATPPSKTKPTAVPKETSAPSEGVTAVPLEIIPSPRKAPKTAAP-----KETPA 1366  
QY 1792 SSSERDRDRDRDREREKRSILTTTVEHAPLWRPGTEQSSGSSGSSGGGSSSRP 1851  
DB 1367 PSPEG-----ATTAPVQIPSPKSGSKA-----GSKETP 1396  
QY 1852 ASHS-----HAHQHSPISPRTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSTSSP 1906  
DB 1397 TTPSEGVTAAPLEIPIISSKTSKMASPKETLVTPSSKKLSQTVGPKETSELEGATAVPLE 1456  
QY 1907 VRPAATFPFPAI---HCPGLGTLGVTLMPEVLLPKAEAPVARPERPRADTGHAFLAK 1962  
DB 1457 IPPSHKKAPKTVDPKQVPLTPSPKDAPTTLAESPSPPSKKAPKTAAPPSEVTT-----V 1510  
QY 1963 PPARSGLPEPASSKSGSEPRPLVPVSGHATIAITPAKNLAPHIASPPPPAPPA--SASD 2020  
DB 1511 PP-----EKPAIPKQASCTASKVPV-----PAETQEVAVSSRETPTVPAVPVKN 1556  
QY 2021 PHREKTOSKPSISQLELRSLRGHSSSYSGEGVPSVSPSSPLTHDKGLPKHLEEDKS 2080  
DB 1557 PSSHKKTSTIELKE-----APATLPP-SPTKSPKIPSSKKAPR-----1594  
QY 2081 HLEGELRPQKPGVKLGGEAAHPLHRLPES-OPSSPLLOTAPGVKGHORVVTLAQHI 2139  
DB 1595 ----TSAPK-----EFPASPSIKPVTTSLAQTAPPSLOKAPS-----1627







Db 762 TSMNSTCTSSAATTVTSSAPSCCTKGRKCSGLMDLYRMLDCSGDYVMSLAKKR 821  
Qy 521 -----DEKEKEAEKEEE-KPEVENDKEDLLKTKTDDTSCEDNDEKEAVAS 566  
Db 822 YRYCGTKRRRFAFKKQLSARKKEKLEVEVEGTGDDGDAEVDGGEAEAEESVS 881  
Qy 567 KGRKTANSQGRKRG---RITRSMAN-----EANSEEAITPOOASLASMELN 610  
Db 882 SSKLKNSPRQPSDAESIRKMLENLPAKRICKTFPVYDNKAKRKNKAKAKSMAKSLV 941  
Qy 611 ESSRWTEEMETAKGLLEH-----GRNWSAIARMVSKTVSOCKNFYFNTYKRONLDE 664  
Db 942 KSKSKSKOORSTKRIYNGHLTRTSRSRSVATSSAAVAAQOORSRKQKQOQK 1001  
Qy 665 ILQOHLKMEKERNARRKKKAPAAASEBAFPVVEDEMEASGSGNEEWEEAEAL 724  
Db 1002 QOQOQOQKQKQAKLEKVPPEP-----PAAVESQMPA----- 1038  
Qy 725 HASGNEVPRGECGPA---TVNNSDTEIPSPHTTEAAKDTGQNGPKPATLGADGPPPG 781  
Db 1039 -----KTEARAPATVTAKSRSAPAEKPSLREL-----PQTL-----PEAA 1075  
Qy 782 PTPPRRTSRAPETPASEATGATPPAPSPSAPPVVPVPEKEEETAAPVVEGE 841  
Db 1076 APNVSVHETE-PASPTPSKQL--MPTPTTPATATKSPAA-----AAVAPPFCAAN 1124  
Qy 842 EOKPP-----AAEELAVDTGKAEPEVKSECT-----EEAEGPAKGDAAEAAT 886  
Db 1125 ROTTNSTVKSRRISDCFAMUTGLKEELKTEQVPPGPGPLQKEKEKFOODRPAEKQKT 1184  
Qy 887 AE-----GALKAKEKGGSGRATTAKSSGAPO---DSDSATCSADEVDEAEGGDN 935  
Db 1185 PQVDHQSPAAVAEKE--VQPKTPKRKAVSRRIKMDATPETLAEPIQER----- 1235  
Qy 936 RLLSPRSLTITGDPRA-----NASPOKPLDLKOLKQRAAAIPIQVTKVHEPPREDA 989  
Db 1236 -----KPVAMPVAVPPVAVAPAILNAAPAPPALOPPFIASMPVAVVVPVPPVPA 1290  
Qy 990 APTK-----PAP-----PPONLOPESDAPQOQSSPRGSRSPAPPADKE 1033  
Db 1291 APLVAAIVPLVAPAPPPVATVAVHPTTRPTTKAAKQMGAPP-----PKPPASLA 1344  
Qy 1034 AFABEAQKLPDPPCWTGSLPPPPPREVIKASPHA-----PPPSAFSAPPQGP 1083  
Db 1345 A-----LKQYPPLEAT--LPVPTNSAPPVAVVPLVPPVPPVPMPLMGQMP 1394  
Qy 1084 LPLGLHDTARPVLP-----RPPTISNPP-----LISSAKHPSVLERQIGATISO-- 1127  
Db 1395 VPVNVN--MLPKPLPLYMTQLQAPANPNPNPAPTAMFVMDHQPLNLTQSQRVLQKPLI 1452  
Qy 1128 -GMSVQLHVPYSEHAKAPGVV-----TMGLPLPMDPKKLAPFGVGVKQEOQLSP 1174  
Db 1453 SGTGDPGLVGLQHRAGMPPARQTCGFERNLIGLDMEMEPLDLKSKSRKPSVPP 1512  
Qy 1175 --RGQAGPPESLGVPTAQASVLRGTALGSVPG-----SITKGIPS 1214  
Db 1513 PRMOQELPMLVPLVLTNANAL--APQSGPGALPTPVGGAPVTCQLPMSGVSGVPG 1569  
Qy 1215 TRVPSDAITYRGSITHGTTPADYLVKGTITRIIGEDSPRLDRGDSLSLPGHVIYEGK 1274  
Db 1570 --VPAPLASHYYSNLDLLKIPQVRNPGTGV-VVPSGPGPAVSAASAAPHPVKSSG-- 1624  
Qy 1275 GHVLSYEGMSVYTC-SKEDGRSSGPPHETAAPKTYTDMGRVGRATISSAIEGLMRA 1334  
Db 1625 GKRRSTEG-----TGSSKDD--NKGQAKACP-RMDEVNHNHIDAINSVIM--AVQAS 1673  
Qy 1335 IPPERHSPHLKQHHIRGSITQGIPIRSYVEAQEDYLRREAKLLKREGTPPPPPSRDLT 1394  
Db 1674 FPDDDEEQAKKE-----KEREREREREKELEEPPLKPSNCI- 1714  
Qy 1395 EAYKTQALGPLKLPKPAHEGLVATVKEAGRSIHEIPREELRHTPELP--LAPRPLKEGSIT 1452  
Db 1715 -----MPLQACAEVLPASVLEKSOAVNAIPAIVAPPVAVVPPVVLAPQPVKPNL- 1763

Qy 1453 OCTPLKYDTGASTTGSKKHVRSLLGS-----PGRTPFPVPHPLDYMDARALERACYEES 1507  
Db 1764 --TPKK-----RSMKSRITIDCRSALLALEETLPAASLPQCIPLPVNGDAK-----EE 1808  
Qy 1508 LKSRRPCTASSGSGIARGAPVIV-PELGKPROSPLTYEDHGAPFAGHLPRGSPVTMREPT 1566  
Db 1809 MFSLP-----DPIPAITPAVVEPQL-----PVQVOGPT 1836  
Qy 1567 PRQGSLSSSKASQODKLTSTPREIAKSPHVTVEHHPHPISPYEHLRLRGVSGVDLYRS 1626  
Db 1837 TSLRDKKTETVMPCNTSTIPTATPLAESPE-----PDPVP-----V 1874  
Qy 1627 HTPPLADPTSIPIRGIPLD-----AAAAYLPRHLAPNTPYPHLYPPYLIRGYPDTALENR 1682  
Db 1875 NLPQA--ETSVPAPVPVSVISVAPVLPPLPTTATPP-----PTTMAETWC 1918  
Qy 1683 QTIINDYITSOQHMHNTATAMAQRAADMLKGLSPRESSLALNYAAGPRGIIDLSQVPHLPV 1742  
Db 1919 SSLMEEH--SSNLNNTSS-----GPHSLAQ-SEQPI 1947  
Qy 1743 LVPPTPGTATMDRLAYLPTAPQPFSSRHSSPLSPGGPHTLTKPTTSSSERERDRDR 1802  
Db 1948 PTAATTAEEAPPVKEDEELPAKKK--QRRRRKNELAAIVADQLLESFKIDNARRONLKL 2005  
Qy 1803 ERDRREREKSLTSTTTVEHAPIWRPGTEQSSGSGSGSSRRSPASHSHAHQSP 1862  
Db 2006 ENLAYEKSEDLTGM-----LLMPSTKRNALGPSAAAAKAKLAKKAADTVA--ESP 2056  
Qy 1863 ISPRTQDALQRPSSVHLNHTMGKIITAVEPSKPT-----VLRSTSTSSVPRPAATFPAT 1917  
Db 2057 ANSPVGRPKRQSCVYRRKAGGVGV--GKATNENISLLKSSLESFSG----- 2105  
Qy 1918 HCPLGGTLDGVYPTLMEPVLLPKAPRVARPERPADTGHAFKAPPARSGLEPASSPK 1977  
Db 2106 -----IEKQLLAKEA-REAKEAQAGGNNAQL--PVSSILRPSILSSA 2146  
Qy 1978 GSEPRPL-----VPPVSGHATIAITPA-----KNLAPHHASP----- 2009  
Db 2147 VAREOQLTRQOOSKLDKQEOQLPPA--ATFSRDPRLNKNHKEQADHKAAPKEQATATT 2203  
Qy 2010 -----DPP-----APPA 2016  
Db 2204 NENPEEDDNYLTEIAKNVNEKIMSATTNEDFEFAHDEFGEDGPDQDQDQDKYVRPPT 2263  
Qy 2017 SAS-----DPH----- 2022  
Db 2264 SMSVRSAPNLNDEHNSFGMCDNNTNTEVMDMDLDEMVSYYTSYSDGLRGGRGRRRRR 2323  
Qy 2023 -----REKTQSKPFSIQELELR-----SLGYHGSY-----SPGVEPV 2056  
Db 2324 RSVLLTRRPKKTORSSELDAEKGCKLCKSPFTATSLSKHNMTLHVSKVSAQEVLSQ 2383  
Qy 2057 SPVSSPSLTHDKGLPKHLEELDKSHLELRLPKQGP-----VKLGEAAH--L 2103  
Db 2384 TAPSPQDGH-----LENKREQAETMLQHEEQEQAQAQAQAQAQAQAQAQAQAQAQA 2437  
Qy 2104 PHLRPLPESQSSSLIQTAPGVKHQVVTLLAQ-----HISEVITQDYTRHHPPQLS 2156  
Db 2438 EKVRQLRVQBOERQOLQREHRRHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2497  
Qy 2157 --APLPAPLYSPG-----ASCPVLDLRRRPSDLXLPDPDGA-PARGSPHSEGGKRSPE 2208  
Db 2498 PTAPTASRLNLPDRLEFECNKLKSAETPR---PHPGGAGAPATSVIYTAGRKQPP 2553  
Qy 2209 PNKTSVLGGEDGIEPVSPPEGMTEPGHRSRVAVYLLYRDGBQTEPSRMGSKSPGNTSOP 2268  
Db 2554 PPPASPSRSRSLPPASPSLPPASPSLPP-----PPA---SPSPSPSP 2601  
Qy 2269 PAFTSKLTESAMVSKKQ 2325  
Db 2602 PA-----EATAPAPVAAAQ-----PAANYCHQPVIOQLFRNPPVPTTTP 2643







Db 1258 ERENAARGADG-----KPSL-TLAATPGPEET-----LTAATVPPPPRRRTAAKSPDLN 1307  
QY 1780 GGPThLTkTtTSS-----SERERDRRE----- 1803  
Db 1308 QNSDQANEWEETASERSSDFASERDGETPPAALMTSKAVGTPGANAGAGPGISAMSRG 1367  
QY 1804 --RDREREKSLTSTTTVEHAPTWRCQESSGSSGSSGGGGSS----- 1848  
Db 1368 DLSQRAKDLKSFSSQRCMDRQNRPGCTGGTSGSSGGGAGPGRTGPGRGDNR 1427  
QY 1849 -----SRPASHSHAHSHSPISRTDALQORPSVLHNTCMKGIIITAVEPSKPTVLR 1900  
Db 1428 SWPSPKNRSRPPPEERPPGLPLPPPPSSSAVERLDQVIH-----SNPAGIOQ 1474  
QY 1901 TSTSSPVREPAATFPATHCPLGTLGVYPTLMEVLLPKAPRVARPERP-RADTGHAF 1959  
Db 1475 ALAQLSSROGNTAPCGH-----PRKPGPPQAPQSSPPPTRYD----- 1515  
QY 1960 LAKPPARSGLPASPSKSGSEPRPLVPVYSGHATIAARTPAKNLAPHASPDPPAPPASAS 2019  
Db 1516 ---PPRAS--SAISSDPHFEEPGWVRGVG-----TPRDSAGVNFPP--- 1554  
QY 2020 DPHREKTKSPFSIQEEL---RSIGYHGSYSYSGVEVPVSPSITHDKGLPKHLEE 2076  
Db 1555 -RRERPPRKPPELLQETVPASHSSGFLGSKPEVPGQESRDSG-----TEALTPHI-- 1606  
QY 2077 LDKSHLEGELREPKQCPVKLGCEAAHLPLPLP-----ESQPS 2115  
Db 1607 WNLHTATSRKSYQPSIE-----PWMEPLSFEDVAGTEMQSOSGVDLSGDSQVS 1658  
QY 2116 SSPLIQ-TAP--GVKGHORVVTLAOHISEVITQDTRHHPOOLSAPLPAPLYSFFCASC 2172  
Db 1659 SGPCSORSSPDGGLKGS-----AEGPKRPGG--PSPLNAVPGESAS 1698  
QY 2173 VLDLRPPSDLYLPDPDHGAPAR-----GSPHSEGGKRSPBPNTSVLGGE 2219  
Db 1699 GSEPEPR--RRPTASHEGERKELPREQPLPPGPIGTERSQRTDRGPEGP----- 1748  
QY 2220 DGIEPVSPPEGTEPG-----HSRSVYPLLYRDE----- 2250  
Db 1749 --LRFAHRPGSOVEFGTNNKSDCLLVGDTLKGKELVASATEAVP-ISRDWELLPSAS 1805  
QY 2251 ---QTEPSRMGS-----KSP-GNTSQPAFFSKLTESNSAMVSKKQKINKLNTNRN 2300  
Db 1806 TSAEPQKSLGSGQCVPSPSGQRPYEVFGSPGPNQV--SGAPIDSQNH----- 1858  
QY 2301 EPEYNISQPGTEIFNWPALTGTGLMTYRSQAOVQEHASTNMGLEAIRKALMCKYQWEE- 2359  
Db 1859 -PNSGGFRPGT-----PSLHQYRSQPLYLPPGPAPPSSALLSGVALKGQFLDFSAL 1907  
QY 2360 -----SPP--LSANAFNP-----LNASASLPAAMPITAADGRSD--H 2392  
Db 1908 QATELGKLPAGGVLYPPPSFLYSAFCPSPLPDPPLQLQVRQLPSP-----SDFYS 1958  
QY 2393 TLTSPGGGKAKVGRPSRK-----AKSPAGLASGRPPSVSVHSEGDCCNRRTPLN 2448  
Db 1959 TPLQPGGSGFLPSGAPQAQMLLPVVDQLPVVNFSLPPA----- 1999  
QY 2449 VWEDRPSSAGSTPPFYNLIMRLQAGVNASPP-----PPGLPAGS-----GPLAG 2494  
Db 2000 -----PPAPPPSLLLPVGPALQPPNLAVRPPPPAPARVLPSPARPFAPSLGR 2047  
QY 2495 HHAWDEPKPLCSOYETLSDS 2516  
Db 2048 AELHPVELKPF--QDYRKLSN 2067



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:38:48 ; Search time 41.42 Seconds  
(without alignments)  
7793.922 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPPNQGFSTEQSRY.....EREPAPLLSAQYETLSDD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5395	42.7	1047	4 Q9NSZ0	Q9nsz0 homo sapien
2	1006	8.0	3469	5 Q9U412	Q9u412 drosophila
3	996.5	7.9	3502	5 Q9VJ39	Q9vj39 drosophila
4	584.5	4.6	5327	5 Q76891	Q76891 drosophila
5	517	4.1	3261	4 Q9Y556	Q9y556 homo sapien
6	507	4.0	3080	5 Q9VRV3	Q9vrV3 drosophila
7	489.5	3.9	2701	4 Q9Y520	Q9y520 homo sapien
8	487.5	3.9	2364	11 Q9ER21	Q9er21 rattus norv
9	485	3.8	5476	5 Q9N117	Q9n117 drosophila
10	485	3.8	5533	5 Q9VPL2	Q9vpl2 drosophila
11	484.5	3.8	3576	11 Q9O2W2	Q9q2w2 mus musculus
12	484.5	3.8	5534	5 Q9NHN1	Q9nhn1 drosophila
13	484.5	3.8	5560	5 Q9VPL1	Q9vpl1 drosophila
14	482	3.8	5533	5 Q9U6C3	Q9u6c3 drosophila
15	480	3.8	3394	4 Q9Y6V0	Q9y6v0 homo sapien
16	473	3.7	2752	4 Q9O035	Q9o035 homo sapien
17	467.5	3.7	2910	10 Q9FND5	Q9fnd5 arabidopsis
18	467.5	3.7	4833	11 Q9QYX6	Q9qyx6 mus musculus
19	467.5	3.7	5038	11 Q9QYX7	Q9qyx7 mus musculus

20	465.5	3.7	2897	5 Q9VID9	Q9vid9 drosophila
21	463.5	3.7	3111	5 Q9VH10	Q9vh10 drosophila
22	462	3.7	4880	11 Q9JLT1	Q9jlt1 rattus norv
23	462	3.7	5085	11 Q9JKS6	Q9jks6 rattus norv
24	460.5	3.6	2768	5 Q9VC00	Q9vc00 drosophila
25	453	3.6	2296	4 Q9UHA8	Q9uha8 homo sapien
26	450.5	3.6	2951	5 Q9W3Z0	Q9w3z0 drosophila
27	450.5	3.6	6815	5 Q9I7U4	Q9i7u4 drosophila
28	449.5	3.6	4900	5 Q9N541	Q9n541 caenorhabdi
29	444	3.5	7962	4 Q10465	Q10465 homo sapien
30	442	3.5	16215	5 Q9NFS3	Q9nfs3 drosophila
31	440	3.5	5170	5 Q17490	Q17490 caenorhabdi
32	438	3.5	6994	5 Q17343	Q17343 caenorhabdi
33	433.5	3.4	1966	5 Q9NHX6	Q9nhx6 drosophila
34	432.5	3.4	3263	5 Q9I7U3	Q9i7u3 drosophila
35	432	3.4	1963	5 Q9VSK5	Q9vsk5 drosophila
36	430	3.4	3257	5 Q9V736	Q9v736 drosophila
37	429	3.4	2649	3 Q9P3U0	Q9p3u0 neurospora
38	429	3.4	6677	5 Q9N435	Q9n435 caenorhabdi
39	425.5	3.4	3851	4 Q43161	Q43161 homo sapien
40	425.5	3.4	3926	4 Q9UPA5	Q9upa5 homo sapien
41	424	3.4	2344	5 Q9N3Y8	Q9n3y8 caenorhabdi
42	415.5	3.3	5120	13 Q9PU36	Q9pu36 gallus gall
43	414.5	3.3	102	4 Q9H4R4	Q9h4r4 homo sapien
44	412	3.3	2829	13 P70039	P70039 xenopus lae
45	409.5	3.2	2361	5 Q94226	Q94226 caenorhabdi

ALIGNMENTS

RESULT 1	
ID Q9NSZ0	PRELIMINARY; PRT; 1047 AA.
AC Q9NSZ0	
DT 01-OCT-2000 (TReMBLrel. 15, Created)	
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)	
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)	
DE HYPOTHETICAL 113.8 KDA PROTEIN (FRAGMENT).	
GN DKEZP434M075.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=TESTIS;	
RA Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;	
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
DR EMBL; ALJ37641; CAB70854.1; -	
KW Hypothetical protein.	
FT NON_TER 1	
SQ SEQUENCE 1047 AA; 113817 MW; B5A1EDE938B7222A CRC64;	

Query Match	42.7%;	Score 5395;	DB 4;	Length 1047;
Best Local Similarity	99.6%;	Pred. No. 1.2e-284;		
Matches 1043;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
Qy 1394	SGQSAIKHNKSLITGTPSKLSRGKMPLEIVPENIKVVERGKYEDVKAGETVRSRHTSYVS	1453		
Db 1	SGQSAIKHNKSLITGTPSKLSRGKMPLEIVPENIKVVERGKYEDVKAGETVRSRHTSYVS	60		
Qy 1454	SGPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTTPPNKST	1513		
Db 61	SGPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTSSNNKST	120		
Qy 1514	NHERKSTLTPTQRESIPAKSPVPGVDVWVSHSPDPHHRGSTAGEVYVSHLPTOLDPAMP	1573		
Db 121	NHERKSTLTPTQRESIPAKSPVPGVDVWVSHSPDPHHRGSTAGEVYVSHLPTOLDPAMP	180		
Qy 1574	FHRALDPAAAYLFQRQLSPPTGPPSYQLYAMENTROTILNDYITTSQOMVNLRPDVAR	1633		















Db 1296 ASAGEATTATGATAAAGVGKGPETATEPAGTAAGADSRPDANDPLAKTASRAINAE 1355  
Qy 656 -----TKSEACKNFYFNYKRHNLDNLLOHKQKT----- 686  
Db 1356 YNATGCGSSSSSNATGASAVQGTGTLNFKPGYQTVVMANVKASTGGDDSGANAGGAAP 1415  
Qy 687 -----SRKPREERDVSOCE-----VAS 704  
Db 1416 GSLAATNASIATSGDKIVKTTTPSRAPNSTSTTAANESSGAGVNTYGHATTAGNYLGO 1475  
Qy 705 TVSAQEDEDIASNEENPEDEVEAVKPSDESPENATSRGNTPEAVELEPT----- 756  
Db 1476 KLKAAQVEGLGAGNELH-----SDVSESKRRKFPENSGAGGAGNATSMNTSSTGSMNIN 1531  
Qy 757 -----TETAPSTPSLAVPTKPAEDSV----- 781  
Db 1532 SHGLKANAKGSMMAKTSMASTSSASVVVTTPSASSSSLSASSMLLSAASVMSTAA 1591  
Qy 782 -TQVNDISISAETAQMD-----VQOEHSABEES----- 809  
Db 1592 ATSSSTATTATASALSPLLDAGSGNSMVNANEITALDGKDKLASCFCVCKAEACPRTRP 1651  
Qy 810 -----VCD----- 812  
Db 1652 LKGRGOQYGPDETIPAGARVCNCOCKSVRSYPNCPPLTCPNPKDRAQLRNIPSL 1711  
Qy 813 -----PPPATK-----ADSDV 823  
Db 1712 FELAPEVRDPLMAFEQIPPHATRCACSACLMIRKLDLPQLNLTGSGGAGSGGDDTD 1771  
Qy 824 V-----EVYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQIN 867  
Db 1772 VTSSECDEREPPGSDTASVESPENLRHKSITWKQOQOQOQOQOQOQOQOQOQOQO 1831  
Qy 868 AORPEPQS-----DN----- 877  
Db 1832 QOPPPAPAOQOQSGSGGDOGTPLIITPRMSSKSGGGAOTAGDNERLLPPAAGQAP 1891  
Qy 878 -----DSSATCSAEDVDGEPEPERMP-----MDKPSLLNPTGSI 914  
Db 1892 KKQKTSEYDSATETADEENENSP-ANRQSPKVLFGHGHGHGHGHANNVAGLPPVANM 1950  
Qy 915 LVSSPLKP-----NPLDLPOLHRAAVIPPMVSCTCPNIPGTPVSGYALYORHIK 965  
Db 1951 GTGGGVOPGGAAGQOVNGPISM-----RREAVNNVQDCV-----FSVIERSLK 1993  
Qy 966 -----AMHESALLEEOROROEI-----DLBCRSST----- 991  
Db 1994 HKGPQPGGQO 2053  
Qy 992 --SPCGTSKSPNREWEVLQAPHQLITNLPBG--VRLPTTRP--TRPPPLIPSSKTTVA 1045  
Db 2054 YRQDPGLKQO 2113  
Qy 1046 SEKPSFIMGISOGTPTGYLT-----SHNOASYTQETPKPSVGSISLGLPROQESAKSA 1100  
Db 2114 -----CAGS-NGTSDSLATSLVNSH-----MGVVGICHPGMAHASSA 2152  
Qy 1101 -----TLPIYIKOBEFFSPRSONSQEGLLVRAQHEGVVGTAGAIQESITRGPT 1151  
Db 2153 GGIGVDKATITPVVKSSGSGSKSGGSASHSTATPTETIYINVPVAHQ-----RGIPPP 2208  
Qy 1152 KISVESIPSLRGSTOQTTPALPQTGI-----PTEALVKGSISRMPEDSSPEKG 1200  
Db 2209 --SQSHVHPAHPSHQ--HPAIPQSHSHQHTQLOQVPEPEPOTLDLSIKKPPRDCGSHPTG 2265  
Qy 1201 REEAASKGHVITYEKGSHILSYDNINKNAREGTRSPRTAHEISLRSKSVESBEGNKQGM 1260  
Db 2266 AGGSSSG-----SGSG-----PSSDRHHGPPPP-----TSMKHIVRS--GMYRGDTV 2310  
Qy 1261 RESPVSALE-----GLICRALPRGSPHSDLKERTV-----LSGSIHQ 1299

Db 2311 TVPSLAAPSSYLTPTRSVKTTIGGGGVVPGVLP-GVPSALYLQVPVVPVPIISG---QG 2366  
Qy 1300 --TPRATTESFEDGLKYPKQIKRESPPIRAFEGAITKPKPDGITTIKEMGRSIEHP-- 1355  
Db 2367 QLPPRA-----GOPPPAQ-----PP-----SGRGVAKVPPK 2392  
Qy 1356 -----RODILTQESRKTPEVVQSTRPIIEGTSISOGTPIKFFNN 1393  
Db 2393 LSPQOAHHLHPHSHGSPSQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2449  
Qy 1394 SGQSAIKHNKSLITGPKSLSRGMPP-----LEIVPNIKVVERGKYEDVAGTVRSR 1447  
Db 2450 AQOQIIVHAPATAAAPSLL---LSPKFDGLVRQTTPEGVSGVGP-----GASGSK 2499  
Qy 1448 HTSVVSSGSPVLSRSLHEAPKAQLSPGIYDDTSARPTVSYQNTMSRGSPPMNRRTSDVTI 1507  
Db 2500 HGSITQGT-----LHMP-----HLESKRPEYKYSQSRHSPAQOQPGNOQL 2544  
Qy 1508 PPNKSTNHERKSTLTPTQRESIPAKSPVPGVDVPSHSPFDPHHRGSTAGEVYWSHLPTQ 1567  
Db 2545 PP-----PQOSS-PQAPPQGYGVGS----- 2566  
Qy 1568 LDPAMPFHRALDPAAYLAFORQLSPTPCYPSQOLYAMENRTQILNDYITSQOM----- 1623  
Db 2567 ---SPYAR-----SPFAGVVEQPQVL---STRQIVMHDYITSQOMQOQ 2604  
Qy 1624 ---OVNLRDPDVARGL-----SPREQ-----PLGLP- 1645  
Db 2605 QO 2664  
Qy 1646 -----YPATRGIDLTNMPPTILVP-----HPGTSTPPMDRITIPGTQ 1685  
Db 2665 YSSRASPADHNSGHDALASFVDVAVQOQLPVPVSKDKSPGSTAPGQ-----VPGSG 2719  
Qy 1686 ITFPPRPYNSMSFGPHTHLAAASAERERERERERERERERERERERERERERERER 1745  
Db 2720 PPLGPSPL-----PPHAVV-----GVAQP 2738  
Qy 1746 GRPGSHGVRSPPSV-RTQETMLQORPSVFGTNGTTSVITPLDPTAOLRIMPLPAGGPS 1804  
Db 2739 PPPTAHDOVRDITLHHHTLVQOOIAQOOHVSRLNVAQVDMOROM----- 2787  
Qy 1805 ISOGLPASRYNTAADAALAYDAASAPQMDVSKTESKHEAA-----RLEENLRSRAA 1859  
Db 2788 -----DOAKRVMRHQHQHVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 2833  
Qy 1860 VSEQOQLEQKTLEYEK-----RSVQCLYT-----SS 1885  
Db 2834 REREREQER 2893  
Qy 1886 AFPSGKPOP-----HSSVYVSEAGDKGPPKPSRYEEELTRGKTT-----ITAANEID 1934  
Db 2894 GAGGGSPGQFLRASV--PETGPPRSIPDRERESYVQAHHGGPAPETPQOLSQSLID 2951  
Qy 1935 VIITQTASDKDA-----RE-----RGSQSDSSSSLSHRYETPSDAIEVI 1976  
Db 2952 AIIKHEINRSDATAGPGRFPSPFVHAPLPGRGSGGGGTGRSS-----PANVLHPM 3006  
Qy 1977 -----SPASSPAPQEKLOTYQPEVVKANQAEENDPTQYEGP 2013  
Db 3007 YLRDLRQPLDGGAGSMLTAENNGKPSVSNIDLDQERISAAAAAAVAAQOQOQOQAP 3066  
Qy 2014 LHYRPOQESPSPQ---QOLPPSOAECMGQVPRT--HRLITLADHICQIITQDFARNOV 2068  
Db 3067 ---PPSOSSRSRSHVQGLRTPTSOAGGSAPSPQOIHFKSITFGELTDSITSDYGTN-- 3120  
Qy 2069 SSQTFQ-QPPTSTFONSALVSTPVRTKTSNRYSPESQAQSVHHQRPGSRVSPENLVDK 2127  
Db 3121 ---PHLRPPPMAYLQETOSIL-PPDRWK-QNRRMQKAEAEAKKH---SQOQOQOQOQOQ 3170  
Qy 2128 SRGRPGKSPERSHVSSEPIPIPPQVPPVHVKQDSLILLSORGAEP--AEQRNDARSP 2185  
Db 3171 QHHAQO 3214











Db	849	SEDERTGGSPVHGSFHDEDDIGSPRLLSYKVGSPKVDKVLDPYSNITVRESLKNP	908
Qy	240	YDENRKK---AEAEKHFIEGPGKPVKVELPLYNQPSDKVYHENIK-----TNQVMRK	287
Db	909	YDSSRRQOMADMA-----KIKLSVLNSDEDLNRWDSQMKODAGRFVFFPNSIIKR	959
Qy	288	KLILFFRRNHARKQBEQIKQRYDQLEMAWEKKVDRTENPNPRKAKESKTRTYEYKQFP	347
Db	960	DSL-----RKRSVRDLDFGEVPSDS---EDGEHK-----SHSPRSALYESRSLFLLRDR	1008
Qy	348	EIRKORQOERFORVGORGA---GLSATIARSEHEISEIDGL---SQEN-----	392
Db	1009	E-DKLREDEDELSSLIERNKFYSFALDKTITPDTKALLERAKSSJSSREENWSFLDWDSR	1067
Qy	393	-----NEKOMRQLSVIP---PMMFDAEBQRKFVINNHGLMEDPMKVYKDRQFMNWTDHE	444
Db	1068	FANPRNNKDEKVDSPAPRIPSWY-MKKKKIR-TDSECKMDDKKEDHKEE-----QER	1119
Qy	445	KEIFKDKFIOHPNFKGLIASYLERKSVPCVLYLYLTTKKNENYKALVRNRYGKR-----	498
Db	1120	QELFASRFLH-----SSEFQ-----DSKKQLHLEKKEEDSDFISGRYIGKOTSEGAN	1167
Qy	499	-----RGRNQIARPSQBEKEEEDKAETKKKEEKDEKDEKEDSK	545
Db	1168	STTDSIQEPVVLFHSRMELTMRQOKEKQPKVEKQEDTENHPKTPESAENKDE	1227
Qy	546	ENTKEXDKIDGTAEETEEREQATPRGRKTANSQGRRKGRITRSMTNEAAAAAASAAAAA	605
Db	1228	--LKTPPSGVPPSVTYTLESAPSALKETGDKTVEAPLVATEEKTVEPATVSEEAKPASE	1285
Qy	606	EPppp-----Lpp-----	613
Db	1286	PAPAPYQLQOVDLPGCADPKDEAAMMPAGVEEGSGDQPPYLDAPPTPGASPSQAESN	1345
Qy	614	-PEPTISTEPTVTSRWTEEMEVAK-----KGLVEHGRNMAAIA-----	651
Db	1346	VDPEPDSTOPLSKPAQKSEANEKPAEKPDATADAEPDANOAKAAPESQSPASEDLEVD	1405
Qy	652	-----KMGVTKSAQCKNFYFNKRRHNDLNLQOHKOKTSKPREED-----	695
Db	1406	PPVAAKDKPKNSKRSKTPVQAAAIVSEKPYTKRSERIDREKLKRSNPRGEAQKLEL	1465
Qy	696	VSOCESVASTVSAQEDDIE-----	715
Db	1466	KMAEKIITRTASKNSAADLEHPEPLSLSTRRRNRVSVYATMGDHENRSVPKEVPQPR	1525
Qy	716	-----ASNEENPEDSEVEAVKPKSED---SPENAT	742
Db	1526	VTRKRLRELOEAAAIVTPTRGRPKPTRRRADEEEENEAKEPATLKKPPEGWNSPKSQ	1585
Qy	743	S-----RGNTEPAVEL---EPTTETAPSTPSLAVPTKPAEDESVEYQVNDISIA	790
Db	1586	TAAGGGPQGGKKGNEPKVDATRPEATTEVGPIGVKKESSMEPKAAEEBAGSEQKDRKDA	1645
Qy	791	ETAEQMDVDQOEHSAEGSVCDPPPATKADSVDVVRVPENHASKVVGCDNTER-----	844
Db	1646	GTDKN-----PPETAPVEVVEKKPAPEKNSKSKGRGRNSRLAYDKS	1687
Qy	845	-DLDRASEKVEPRD-----EDLWAAQINQAPE--PQSDNDSSATCSADE--DVDGEP	893
Db	1688	ASLKNVDAVSPRGAQAAGERESGVVA--VSPKESPEQKEDGLSSQLKSDVPDPDPKEP	1745
Qy	894	ERORMPMDSKPSL-----LNPSTGILVSSP--LKPNDLDPQL	930
Db	1746	EKEDVASGSPSEATOLAKOMELEQAVEHIAKLAESASAAKADAPGLAPEDROKPAH	1805
Qy	931	Q-----HRAAVIPMW---VSCPTCNPIPTGPVSGVALYQRIKAMHESALLEEQRQREQI	983
Db	1806	QASETELAAAGSIINDISGEFENPAPPPYPG-----ESQT	1842
Qy	984	DLERCSTSPCGTSKSPNREWELQAPHQHILNPEGVRLPT---TRPTRRP-----PPLI	1037











Db 3063 RAALGMPKPYDINR 3076

## RESULT 7

Q9Y520  
ID Q9Y520 PRELIMINARY; PRT; 2701 AA.  
AC Q9Y520;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 295.8 KDA PROTEIN.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rhodes S., Huckle E.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL096857; CAB51071.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 2701 AA; 295828 MW; D2043FF9840ADECD CRC64;

Query Match 3.9%; Score 489.5; DB 4; Length 2701;  
Best Local Similarity 19.2%; Pred. No. 7.5e-18;  
Matches 514; Conservative 338; Mismatches 910; Indels 915; Gaps 125;

QY 6 YPPNQAFSTEQSRYPHSHVQYTPFTRHQ---QEFVAPDYRSSHLEVSQASQL----- 56  
Db 258 YPPLHGP-----RFP-SISETNKLGRGPPPSWASEPERPSILSASELKELDKFDNL 311  
QY 57 -----LQOQOQOQLRRRPSLLSEPHGSDRPOERTSYEPFHPGSPVDHDSLESKR 108  
Db 312 DADEGWAGQMEVDYTEQLNFSDDEQGSNKPENNE-----DQGSKASEN 360  
QY 109 PRLQVSDSHFORVSAAVLP-----VHPLPEGLRASADAKKDPAF-----GKHEAPSSP-- 159  
Db 361 NENKKTDEVSTNKSQIQAQPSVAKVPYG-----KGPSFNQERGTSSHLPPPKL 412  
QY 160 ISGQCDDQNASPSK-----LSKEELQSDVDRETAKVEQOILKLLKKQOQ-----LEE 211  
Db 413 LAQHQPPDQAVGPRGPPSPKQOAV-----EDEIWKORRRQOQSEISAIVE 460  
QY 212 EAAKPPPEPEKPVSPVPPVQKHRSIVQIYDENRKK---AEEAHKIFEGLGPKVPLXNOP 269  
Db 461 RARRE-----EERR-----MEQRKAACAELKRLDEKLG-----ILKQP 499  
QY 270 SDTKVYHENIKTNQVMKKLILFFKRRNHARKQEQKICQRYDQLMBAWEKKYDRIENNP 329  
Db 500 SPEI-----RERERERER-----EKELEKEQE 526  
QY 330 RRAKESKTRREYKQPEIRKQEQERQVQRCQAGLSATTARSEHEISEIIDLGE 389  
Db 527 REKERE-KDREOQEKELKEKEQEKREMEK-----ERQEKERE-----LER 569  
QY 390 QENNEKQMRQLSVIPPMFDAQRRVKFINNGLMEDPMKYKDRQFMVNTDHEKEIFK 449  
Db 570 QKEKELELQMK-----EKEKELELQMK-----EKEKELELQMK 589  
QY 450 DKFTQHPKFNGLTASYLERKSVDPCLVLYLTKKNENYKALVRRNYKRRGRNQIARPS 509  
Db 590 EK-----EREKL-----EREKL-----EREKL-----EREKL 596  
QY 510 QEKVEKEEDKAEKTKKEE-----KKDEBEKDEKSKENTKDKIDGTAETEREQ 566  
Db 597 -EKEIEPREPNLEPMWKEQSENKSCNEEPEVFTRODSNRSEKATPV---VHETEPESG 652  
QY 567 ATPRG-----RKTANSGRKKGRITRSMTEFNAAAAASAAAA 601  
Db 653 SOPRAVLGYFKQFKSLPRFORQEQEMKQOQOQOQOQOQVLPQTVSOPSSSTV--- 709  
QY 602 AATEEPPPLPPPPPEIPTEPVSRTWTEEMEYAKKGLVHGHRNAAIAKMWGTSKSEAQ 661

Db 710 -----PPPHRPLYPMPQHP-----QHLASMGF-----DPRWLMQSQVMDPRM--- 748  
QY 662 CKNYFYNYKRHNLDNLQHQKQTSRKPREEROVSQCEASVTSVAQEDDEIEASNEE 721  
Db 749 -----MSGRPAMD-IPPIHGMIPPKPLMRD-----OMEGSPNSSEFHIARSADH 796  
QY 722 NPDESEAVKPSDESPENATSRGNTEPAVELEP---TTETAPSTSPSLAVPSTK----- 773  
Db 797 AISISEPRLWGSDBPYPHAEQQOATTPKATE-EPEDVSEALDOEQITAAYSVEHNOLE 855  
QY 774 --PAED---ESVETQVNDISIAETAQMDVDQOQEHSAEBSVCDPPPPATKADSDVVEYRV 828  
Db 856 AHPKADFTRESSEAQVQKLSRVE---DVRPHHTDANNQACPEAPQKLTSLAPQEEIRI 912  
QY 829 PENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQOQINAOBPPEQSDNDS-SATCSADE 887  
Db 913 -----SAVESQPSRKRVSHGSHNTQKPD-----ORSEPSAGIPKVTSCIDSK 957  
QY 888 DVDGEPEQRMPMDSKPSLLNPTGSIILVSSPLKPNPLDLPLQLOHRAAVIPPM-----VS 942  
Db 958 EPIERPE-----EKP-----KKEGFIRSESGPKPEKVKSKSETRWGPRFSSNRREVN 1006  
QY 943 CTPC--NIPICTPVSGYALYQRIKAMHESALLERQORQEQIDLECRSSTSPCSTKSP 1000  
Db 1007 DRPVRRSGPIKKPV-----LRDMK-----EEREQKEKEGEKAKEKVTKEV-VVRPE 1051  
QY 1001 NREVEVLOPAPHQLITNLPEGVRLPTTRTPP-----PPPL-----IPSKTTVASEK 1048  
Db 1052 KTEKDKLPPPP-----PPQPPAPIQPSVPPPIQPEAEKFPSTETATLAQK 1098  
QY 1049 PSFIMGSGISQGTGCTVLTSHNOASYTQETPK-----SVGSIISGLP--- 1091  
Db 1099 PS-----ODTEKPLEPVSTVQVEPAVKTVNNOOTMAAPVVK 1133  
QY 1092 --ROQESAKSATL-----PYIKQEEFSP-----RSQNSQPE----- 1120  
Db 1134 EEKQPEKVISKDLVIERPPRPSRPAVKKESTLPRTYWKARERDWFPDQVYGRGRGEY 1193  
QY 1121 ---GLLVRAQHEGVVGTAGAIQESITRGTP---TSKISVESIPS----- 1160  
Db 1194 YSRGRSYGSGYGRGRGRGH-----TRDYPQYRDNKPRAEHIPSGPLRRESETRSE 1247  
QY 1161 -----LRGSITQGTTPALPOTGIPTEALVKGSGISRMP---IEDSSPEKGRGEEAAS 1206  
Db 1248 SSDEEVVPKRRRQSGSETDSEIHESSADKSLKGLPKREERPNKKPVKPHISSFKP 1307  
QY 1207 KGHVYE-----GKSGHLSYDNINKNAREGT-----RSP---RTAHEISLKR 1245  
Db 1308 DNHVRIIDNRLLEKPVVDDDKAKPGFLPKGEPTRRRGGTFRRGDRDPPGRPSRPTLRR 1367  
QY 1246 -SYESVEGNIKQGMRESVPSAPLEGILICALRPGSPHSLDKERTVLGSGIMOGTTPAT 1304  
Db 1368 PAYRDNQWN-----PRQSEVPKPEDGEPRRHEQFPIAADKR-----PPKF 1409  
QY 1305 TESPEDEGLKVPQIKRESPP-----IRAFEGATTKGPKYDGIITIKEMGRSIEHI 1354  
Db 1410 ERKEDPARERPRRQRTPPRQDKPPFRRLREREA- SKSNEVVAVPT-----NGTVNVV 1464  
QY 1355 PRODILTQESRKTPEVVQSTRIIEGSGIS-OGTPIKFDNNNSGQSAIKHNKSLITGPSKL 1413  
Db 1465 AOEVPNTL-----GDISGNKTPDLSNQNSDDQANEE----- 1495  
QY 1414 SRGMPPLIIVPENIKVVERKYEVDVKAQETVRSRHTSVVSSGSPSVLSTLHAPKAQLSP 1473  
Db 1496 -----WETASESSDFNERRERDEKKNADL---NAQTVKVGENVL-----PPKREIAK 1540  
QY 1474 GIYDDTISARTPTPVSYQNTMSRGSPMNRSTSDVTIPPNKSTN-----HERKSTLTPTQRES 1528  
Db 1541 RSF---SSQR-PVDRQNRNGNG-----PPKSGRNFSGPRNERS-----G 1577  
QY 1529 IPAKSPVPGVDVPPVSHSPFDPHHRGSTAGEYVWHLPTQLDPAAMPFHRALDPAALVQ 1588







Db 724 RSLMSS--PEDLTDFEELKAEEDVARDIKPOLELIEDEEKLRETEPGEAYVQKETE 781  
Qy 758 ETAPSTSPSLAVPSTK-----PAEDSVETQVNDSI----- 788  
Db 782 SKGSAEPGDEGITTTEGECEQTEPELEPVKEGVDDIEKFEBEGAGFESSAGYEE 841  
Qy 789 --SAETAEDQMDVDOOQERHSAEGSYCDPPATKADSVDEVRVPHENHASKVBGDMTKERDL 846  
Db 842 KAETEEAEPEDEEDNVSASAKHSPTDEDEIAKAEADVHIKRESVASDRAEEDM 901  
Qy 847 DRASEKVEPRDEDLVAAQINQAQRPQSDNDSSATCSADEVDGCEPERQRMFMDSKPS 906  
Db 902 DEALEKGE-----AEQSEGEDEEEDKAEDAREE--DHEPDKTE--AEDYVMA 945  
Qy 907 LLNPTGSLVSS--PLKPNPLDLPOLQHRAAVIPPMVSCPTCPNIPCTPVSGVYALYQRH 963  
Db 946 VVDAEAGVTEQDYFLGTAPAKQGVQS-----PSREPAS----- 981  
Qy 964 IKAMHESALLEEQROQFQIDLECRSSTPCGTSKSPNREWEVLQAPAPHQLITNLPEGVR 1023  
Db 982 --SIHDETL-----PGGSESEATADEENRE--DQPEFTATSGYTOSTI 1022  
Qy 1024 LPTRPT-----RPPPLIPSSKTTIVASEKPS--FINGGSIISQGTGTYLTHSHNOASYTO 1076  
Db 1023 EISSEPTMDMSTPRVMTDETNEETSPSQEFV-----NITKYESSLYSQ 1070  
Qy 1077 ETPKPSVGSISLGLP-----RQGESAKSATLP--YIKQEEFSPRSONSQPEGILV 1124  
Db 1071 EYKPVVASFN--GLSDGSKTDATDGRYNASASTISPPSMEEDKFKSALRD-----AY 1124  
Qy 1125 RAQHEGVVGTAGAIQSGITRGTPTTSKISVESIP-----SLRGSITQGTPAL 1172  
Db 1125 RPEETDVKTAGAELDKOVSDERLSPAKSPSLSPSPSPPIEKTPLGERSVNFSLTPNEIKA 1184  
Qy 1173 POTGIPTAALVKGISIRMPED--SSPEKGREEAASGHVYIEKSGHILSYD---NIKN 1227  
Db 1185 SAEGEAT--AVVSPGVQAVVEHCASPEEKTLEVYVSPQSV--TGSAGHTPYIQSPTEKS 1242  
Qy 1228 AREGTRSPRAHETSILKRSV--ESVEGNIKQGMMSRQSVAPLEGLICRALPRGSPHSDL 1286  
Db 1243 SHLPEVTENAAQVPVFETEAKDENERSISPDPEV-----POSES 1286  
Qy 1287 KERTVLGSGIMQGTPRATTES--FEDGLKYPKQI---KRESPPIRAFEGAITKGPYDGIT 1342  
Db 1287 PIEKVLSP--LRSPPLIGSEAYEDFLSADDKALGRRESF-----FEGKNGKQGFSDKES 1340  
Qy 1343 TIKEMGRSIEH-----IP--RQDILQESKRTPEV--QSTRILIEGSI--SQGTPI 1388  
Db 1341 PVSOLTSOLYQDKQEKERAGFIPIKEDFSPKAKASDAEIMSSQSALALDERKLGSGSPT 1400  
Qy 1389 KFDNNSGSAIKHNKVSILITGPKSLRGMPPLEIVPENIKVVERKVEDVKAGETVRSRH 1448  
Db 1401 QVD--VSQFGFKEDTKMSISEGTVSDKSATP-----VDEGAEDTYSHMGCVASVS 1449  
Qy 1449 TSVVSSGSPVLRSTLHBAKQAQSLPGIY-----DDT-----SARRTPVSYQNTMSR 1494  
Db 1450 TA-----SVATSSFEPTTDDVSPSLHAEVSGSPHSTVEDDLSVSVVQTPPTTFOET --- 1500  
Qy 1495 GSPMKNRTSDVTIPPNKSTNHERSTILPTQRESIPAKSPVPGVDPVVSHPFDPHGRS 1554  
Db 1501 -----EMSPSKE--ECPRPMSTSPDFSPKTKARTPVOD-----HRSE 1537  
Qy 1555 TAGEVYVWHLPTQLDPAFPHR--ALDPAAAAAYLFORLSLPTPGYPS--QVOLYAMENTRQ 1611  
Db 1538 Q-----SSMSIEPQESPEHSLAND-----FSRQ---SPDHTVCGAGMLHITENGPT 1581  
Qy 1612 TILNDYITSOOMQVNLPRDVARGLSPREQLGLPYPATRGIDITLNNPPTILVPHPGGTS 1671  
Db 1582 EV--DYSPPDIQDSSL---SHKIPPTPEP---SYTDNDLSELISVQSEASPTSSAH 1632  
Qy 1672 TPPMDRITYIPGTQITPPPPRYNSASMSPGPHILAAASAEERERERERERERERERERER 1731  
Db 1633 TP-----SQIASPQEDTLSDYVVPVPRDMSLYASLASAEKVQSLGEKLSKPSDISP 1682

Qy 1732 ASSDLYLRLPGSEQPRGSGHYVRSPSVRTQETMLQARPSVFQGTNGTSVITPLDPTA 1791  
Db 1683 -----LTPRESSP-----TYSFGFSDTSIGAKESTAAAYQ-----TSSPPPIDAA 1722  
Qy 1792 QLRIMPLPAGGSPISQGLPASRYNTAADALAALDAASAPQMD-----V 1836  
Db 1723 -----AEPYGFRS-----SMLFDTMQHHLALSRLDTTSSVEKONGGKTPGDFNAYQKP 1771  
Qy 1837 SKTKES-----KHFAA-----RLEENLRSAASAVSEOOQLEQKLEVEK 1875  
Db 1772 ESTTESPEDEEDYDYESHEKTIQAHDVGYYEYKERTIKSPCDSGSYSETTEKTKTPED 1831  
Qy 1876 RSVQCLTSSAPFGKPOPHSHSVYVSEAGDKGPPKRYEEELTRGKTKTITAAANFIDV 1935  
Db 1832 GGYCEITEKTRTPPEGGYS---YEISEKTRTPPEVSGYTYEKTERRRLD----- 1881  
Qy 1936 IITROIASDKDAREGSSQSDSSSSLSHRYET-----PSDAIEVISPASSAPPQE 1987  
Db 1882 ---DISNGYDDTDGGHTLGD---CSYSYETTEKITSPFSESESYSYETTTKTRSPDT 1933  
Qy 1988 KLQIYQ--PEVVKANQANDPTROYEGLPHHYRQOESPS-----PQ 2027  
Db 1934 SAYCYETMEKITTKTPOAS---TYSYETSDRCYTPERKSPSEARQDVLCLVSSCEFKHPK 1990  
Qy 2028 QOLPPS---SOAEGM--GOVPRTHRLITLADHICQIITODFAR-----NOV 2068  
Db 1991 TELSPFINPNPLWFAGEEP-----TEESERPLTQSGGAPPSPGGKQ 2034  
Qy 2069 SSOTPOQPTSTFQNSALYSTVTRTKTSNRYSPESQAQSVHHQRPGRSVSPENL--VD 2126  
Db 2035 GRQDETPTPTSVSEAPS-----QTDSDVPPETE-----ECP--SITADANLSED 2078  
Qy 2127 KRSRGRCKSPERSHVS--SEPYEPISP--POVPVVEHQDSLLLSORGA---EPARQR 2179  
Db 2079 ESETIPTDKTVYKHMOPPPAPMQDRSPRHPDVSMVDPEALAEQNLKALKDLKEK 2138  
Qy 2180 NDARSPGISYLSPFKLENTSPMKVSKQEIFRKLNSGGSDMAAQAOPGTEIFNLP 2239  
Db 2139 AKTKPG-----IKTKSSPVKK-----GDGSKSPSAASP-----KPG 2171  
Qy 2240 AVTTSGSVSSRGHGFADPASNLGLEDIIR-----KALMGSDDKVEDGHGVMSQPMGV 2292  
Db 2172 ALKESSDKVS---VASPKKESVEKAMKTTTPEVKATRGEEKDKETKNAANASAKSV 2228  
Qy 2293 V-----PCTANTS 2300  
Db 2229 KTATAGPGTTKTA 2241

RESULT 9

Q9NJ17

ID Q9NJ17 PRELIMINARY; PRT; 5476 AA.

AC Q9NJ17;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE SPLIT ENDS.

GN SPEN OR CG18497.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20157049; PubMed=10655223;

RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,

RA Suh C., Voas M., Williams A., Rubin G.M.

RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated

RT Protein Kinase Signaling Pathway That Interact With the yan Gene of

RT Drosophila Identifies split ends, a New RNA Recognition Motif-

RT Containing Protein."















QY 1873 -----VEKRSVQCLYTSAFSGKPPQHPSSVVS----- 1901  
DB 4657 MPQHPKKAIAAAETESYQAI--NSSIPSGGLPMHOTAAATQKITGGVADAVSKALVDP 4714  
QY 1902 -----EACKDKG-----PPKRSYEEELTRGKTTITAAAN- 1931  
DB 4715 VTGVITAGMPCKEGNGLPAATAAPANSNEDGQAAPPLQHOHQOQOHPQPOQOQANL 4774  
QY 1932 -FIDVIIITQIASOKDAREGSDSSSSLSHRYETPSDAIEVISPASSPAPPOEKILQ 1990  
DB 4775 QINTLIPSLGNPITALGKSVQLETSAAALL-----NKPVSVLVKGNA-----SQVIQ 4823  
QY 1991 TYQPEVVKANAENDPTRO-----YRPOQ-----ESPPOQO 2029  
DB 4824 OQOQOIVA-----PAKQPIIQQNPLPTVLHHAQHTVVRPQPKAHLNREKNIQOQ 4876  
QY 2030 LPPSSOAGSGMGOVTRHLITLADHICOIITQDFARNQ-----VSSQTPQO----- 2075  
DB 4877 LPTTKQAVAQ-----PPQH-----APHSGHMLITDTAGNQQOLVOPQIIARHLOQQOHLQVNV 4928  
QY 2076 -PPTSTFQNSPALSALVSTVTRTKTSNRYSPESQAQSVHQRPGSRVS-----PENLVDKSR 2129  
DB 4929 PPPTAHSPPHSRI-----PSOQOQ-----LGFQASISPOOQOQOQPTVVIKQA 4969  
QY 2130 GSRPGKSPERSHVSSEPEYIPSPQVPVHKEQDSLILLSQSG-----AEPAEQRNDARS 2184  
DB 4970 AS--AAQOQILHVSSKASVVPQO-----QQQLPPTSTSGPHLQAKP----- 5011  
QY 2185 PGSTSYLSFFTKLENTSPMKSKKOE-IFRKLNSGGGSDM-----AAAPGCTEI-F 2236  
DB 5012 --NYSYAP---TVLTPTLPVAVQOQOQOHLKYKONNQKQGAQIOMPHGIMPTHPGMLLQO 5066  
QY 2237 NLPAVTTSGSVSSRGHSPADPASNLGLDIIIRKALMGSDFKVEDHGVVMSQPMGVVPGT 2296  
DB 5067 KLPA-----HLOQOQOHL-NPSPPGKPNVL-----HGLQSGQ---IMFGS 5104  
QY 2297 ANTSVVTSGTRREEDGSPHSGGVCK-----PKLISKSNRSRKS 2337  
DB 5105 V-----GSPPPVSAVLKTAQQQVNSVVPVAGIRTAIPNISQSPQSRVSP 5150  
QY 2338 PIPQOGLGTE-----RPSVSVSHSGDY-----HROTPGWAMEDRPSSTGSGTFPNP 2387  
DB 5151 VLP-PGISGVPPFDASLNDSTYRGVTASRDFMLYQHHLMRGDDYDKMGSS-----PP 5202  
QY 2388 LTRMLSLSTPTPTACAPSVAQNQAAPHQONRIWEREPAPLLSAOYETLSDS 2438  
DB 5203 LELRRPGSGPPTI-----AVPHS-----LSPQORTAADS 5233

RESULT 11  
Q9Q2W2  
AC Q9Q2W2 PRELIMINARY; PRT: 3576 AA.  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE MSX2 INTERACTING NUCLEAR TARGET PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP TISSUE=TESTIS;  
RC  
RA MEDLINE=9379811; PubMed=10451362;  
RA Newberry E.P., Latifi T., Towler D.A.;  
RT "The RRM domain of MINT, a novel msx2 binding protein, recognizes and  
RT regulates the rat osteocalcin promoter."  
RL Biochemistry 38:10678-10690(1999).  
DR EMBL; AF156529; AAD55931.1; -  
DR InterPro; IPR000504; -  
DR InterPro; IPR001993; -

DR Pfam; PF00076; Irm; 3.  
DR PROSITE; PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
DR SMART; SM00360; RRM; 1.  
SQ SEQUENCE 3576 AA; 390966 MW; 9B56855D8A8F38A8 CRC64;

Query Match 3.8%; Score 484.5; DB 11; Length 3576;  
Best Local Similarity 19.9%; Pred. NO. 2e-17;  
Matches 492; Conservative 298; Mismatches 875; Indels 811; Gaps 118;

QY 4 SCYGP-----NOGAFSTEQSRYPHPSVQYTFNTRHQEFAYPDYRSSHLEVSQASQ 55  
DB 1500 SDFPPGRLYGQAEGANSTSDSVOEP-----VVLFRHSRMELTR----- 1539  
QY 56 LIQOQOQOQLRRRPSLLSEFHFGSDRPOERRTSYEP-FHP-GPSPVDHDSLSKSKPRLEQ 113  
DB 1540 -MQQKEKEK-----DQKPEAEKQEEPEHPTKTEP-----AAETKEPEPKA 1580  
QY 114 VSDSHFQRYSAVLPLVHLPLPEGLRASADAKDPAGFGKHEAPSSPISGQ----- 163  
DB 1581 PVSAGLPATITVV-TPEPASSAPEKAEAEAEAPSPAGEKPAEPAPVSEETKLVSPEVSV 1639  
QY 164 -----PGDDONAS-----PSKLSKEELIQSDRDVRE-----IAKVE--- 196  
DB 1640 PVEQPRQSDVPPGEDSRDSQSAALAPSAQESAATDAVPCVNAEPLTPGTTVSQVSSV 1699  
QY 197 -----QOILKKKKQOOLEE-----EAAKPPPEPEKPVSP 226  
DB 1700 DPKSSPQPLSKLTQSEAEKGVEKPDTPSTPDTATQNGAVASEAQPPASEDEANP 1759  
QY 227 PVEQKHSRIVQIYDENRKAAEAHKIFEGLGPKVVELPLYNQPSDTKYVHENIKTNQVMR 286  
DB 1760 PVAADR-----KTNKSRKSTSVQAAAAVVEKPVTR-----KSERIDR 1799  
QY 287 KKLILFFRRNHARKQREKICQRYDQLMEAWKKVDRLENN-----PRRKA 333  
DB 1800 EKLKRSSSPRGEAQKLELK-----MEA--EKITRASKSSGGDTEHPPEPLPLSRS 1849  
QY 334 KESKTREY-----EKOPPEIRKQEQOERFORVGORGAGLSATIAKSEHSEIIDL 387  
DB 1850 RRRNVRSVYATWDHESRP-AKEPVEQPVTR-----KRLERLEQ----- 1890  
QY 388 SEQENNEKQRLSVIIPPMFMDAEQRRYKFINMGLMEDPMKYKDRQFMNVYVWTDHEKEI 447  
DB 1891 -----AVVP-----TTPRRGR-----PP-----KTRRAEEDGEHERK- 1919  
QY 448 FKDXFIQHPKNEGLIASYLERKSVDCVLYYYLTAKNYKALVRRNYGKRGRRNQOIA- 506  
DB 1920 -----EPAETP-----RPAEGWRSPRSQKSAA 1941  
QY 507 -----RPSOEKVEKEEDKAEKTEKKEEK-----KDEEKEDEKSKENTKDKIDG 556  
DB 1942 AAGPOGKGRNEQKVEAAAAAGAAQASTREGNPKSGEREAAASEPKRRDRDSTDSGPD 2001  
QY 557 TAEETEEREQATPRGRTANSQGRKGRITRSMTNEAA-----AASAAAAATEEP 609  
DB 2002 FPVEVLERKPP-----EKTYKSK-RGRARSTRSGMDRAAHQHSLEMAARAAGQAADKEAGP 2056  
QY 610 PLPPPE-----PISTEPVETSRWTEEMEYAKGLVHEGRNWAIAIKMVGTK 657  
DB 2057 AAASPOESESQKSGSSPQLANNPADPDRAEESAS-----ASTAPPEGTQ 2104  
QY 658 SEACKNFFYKRRHNDLNLQHQKTSRKPREERDVDSQESVASTVSAQEDEDIAS 717  
DB 2105 LARQIE-----LEQAVQNIAPL--EPSAAAASKGTATATA-----TAA 2141  
QY 718 NEENPEPSEYAVKPSDESPENATSRGNTPEPAVELETTETPTSTPSLAVPSTKPAED 777  
DB 2142 SEEPAPHGHPKPAQASE--TELAAGISIIISDASGEFENFSAFSPV-----PGSQT 2195  
QY 778 ESVETQVNDSTISA--ETAEQMDVDQOQHSAGEGVSVDPPPA-----TKASVDVVEVRVP 829  
DB 2196 EGMEPGLHEAESGILETC-----TATESAPQVSALDPPPEGASDTRKTRGNSG---SVQ 2247



QY 830 ENHASKVEDNTKERDLDRASEKVEPR-----DEDLVAAQIINAQRPQSDNDSDSATCS 884  
DB 2248 EAKGSKAEVTPRK---DKGRQKTRRRKRANKKVAITETRASEAE-QQSESPA---2300  
QY 885 ADEDVDGPERQRMFMDKPSLLNPT--GSILVSSPLKPNLDLPQLQRAAIVPPMYSC 943  
DB 2301 AEEATAATPEA---POEEKPSEKPPSPAECTFDPKTSKTPPAESLSQENSA-----EK 2350  
QY 944 TPCNPIPTGTPVSGYALYORHKKAMHESALLEEQRORQOIID--LECRSSTPCGTSKSPN 1001  
DB 2351 TPCKAPV-LPA-----LPLSQPALMDGQPAQREKVHSHIESDPVTPPDSGIPP-2399  
QY 1002 REWEVLQAPAHOLITNLPEGVRLPTRTRPPPLI-----PSSKT-----T 1043  
DB 2400 -----PTIPLVTIAKLPPV-IPGVPHQSPPPKVTWITROEPEPRAQSTPSPALPDT 2452  
QY 1044 VASE-----KPSFMGSGISOGTPTGYLTLSHNOASYTOETPKPSVSGISLGL 1090  
DB 2453 KASMDTSSSTLRKILMDPKYVSATGVTSTVTTAIAEPVSAQCILQEAAPPCD-----2506  
QY 1091 PROQ--ESAKSATLPYIKOEESPPSONSQPEGLLVRAQHECVRGRTAGATQEGSITRGT 1148  
DB 2507 PKHPPLEGVSAAVP-----NADTQASEVPVAADKEKVAPVIAPKIT--SVISRM 2554  
QY 1149 PTSKISVESIPSLRGSITQGTALPO--TGIPTEALVKGSISRMPIED--SSPEKGREAA 1205  
DB 2555 PVS-IDLEN---SOKITLAKPA-PQTLTGLVSAITGLVNLVLPVNAKGPVGK-SVAT 2607  
QY 1206 SKGHVYIEGKSGHILSYDNINAREGTRSPRTAHEISLKRSYESVEGNKOGMSRESVP 1265  
DB 2608 LKG--LVSTPAGPVNLLKGPVNLTPGVNLTTPVSATGVTGNAAPGPVTAACGVTATTG 2665  
QY 1266 SAPLEGLICRALPGSPSHDLKERIVL---SGSIMQGTPTATTESFEDGLKYPKQIKRES 1322  
DB 2666 TAAVTGAVTAPAAKQKQASSENRSFRHPGMSVIDDRPADTGG--AGLRVNTS---EG 2720  
QY 1323 PPIRAFEAGITKGPYDITTIKEMGRSIIHEIP-----RODILTQSRKTPPEV-VQSTRPII 1378  
DB 2721 VVLLSYSGQKEGP-----QRISAKISQIPPASAMDIIEFQQSVKSQVKAADSIPTQ 2772  
QY 1379 EGSISQGTPIKFDNNSGOSATKHNVKSLITGP--SKLSRGMPPLIIVPENIKVVERGYE 1436  
DB 2773 SAPKGPQTPSAFANVAHASTLVLTAQTYNASPVISSTVDRPSLE-KPEPIHL-----SVST 2828  
QY 1437 DVKAGETVRSRHTSVVSSG---PSVLRSTLHEAPKAQLSPGIDYDTSARRTPVSVQNTMS 1493  
DB 2829 PVTQGTVK-----VLTQGIWTPPVV---VHN--QLVLTPIVITNKKLADPVT-----2872  
QY 1494 RGSPMMNRTSDVTIPPNKSTNHERKSTLTPTQRESIPAKSPVPGVDPVYVSHSPDFPHRG 1553  
DB 2873 -----LKTIETKVLPANLG-----PTLTHPPALPSKLPAE-----VNHVPSGP---2912  
QY 1554 STAGEVYVSHLPTQLDPAMPFHRALDPAAAYLFORQLSPITPG-----YPSQVLYAME 1607  
DB 2913 STPADRTIAHLAT---PKPDTHS-----PRPTGTPGLFPRCPHSS-----2951  
QY 1608 NTRQITLNDYITSOQMVLNRPDVARGLSPREQLGLPYPATRGIIDLITNMPPIILVPH 1667  
DB 2952 -----TISTALSTNATVNLAAAGI-PVQFISSIHPEQSVIMPHPSITQVSLGHL 3000  
QY 1668 G-----GTSTPPMDRITYTPGTITFPFPRYNASMGHPHPLAAASAEEREREREKE 1723  
DB 3001 SOGEVRNSTPLPSTISYIRETLHSR-----APLQPOOIEARAPO-----3042  
QY 1724 RERERIAAASDLVRLGSEQOPGRPGSHGVYRSPSPSVRTQETMLQORPSVFOGTNGTSV 1783  
DB 3043 -----RVGTQPOATTG-----VPALATQHPPEEE-----3066  
QY 1784 ITPLDPTAQLRIMPLPAGGPPSISOGLPASRYNTAADALAAALVDAASAPQMDVSKTESK 1843  
DB 3067 -----VHYHLPVAR-----AAAP-----3079

QY 1844 HEAALEENLRSGAAVSEQQOLEQKLTLEVEKRSVQCLYTSSAPSGKPOPHSSVYVSEA 1903  
DB 3080 -----VQSEVLVMOSEYRLHPYTV---PROVRIMV-----HPHVTAV-----3113  
QY 1904 GKDKGPPPKSRYEEELTRGKTITTAANFIDVITITROIASDKDARERGSOSDSSSLSS 1963  
DB 3114 -----SEQPRATEGVVVPANKAPQLVKEAVKTSDAK-----3147  
QY 1964 HRYETPSDA-LEVISPASSAPP---QEKLOTYQEVVVKANQAENDPTROIEGPHLHYRP 2019  
DB 3148 ---AVPAPAPVFPVFPVTPAPPHPGEARILT-----3176  
QY 2020 QOESPSQOQLPPSSQAEQMGQVPRTHRLITLADHICQII-----TODFARNOVSS-QT 2072  
DB 3177 -----VTPSQLOGLPLTPP-----VVVTHGVOIVHSSGELFOEYRYGDRVTHA 3221  
QY 2073 PQQPPTSTFQNSPALVSTPVRTKTSNNKSPESQ-AQSVHHQRPQ-----SRVSPENLV 2125  
DB 3222 PAOOLTHT-QPPVASSISLASRTKTSQAQVPEGEPLQTSQAQAPAPSTQATQPIPPAPPC 3280  
QY 2126 DKSRGSRPKGKSPERS--HVSE-----PYEPISPPQVVPVHEKQDSLILLS 2169  
DB 3281 QPSOLSQAQPPSKIPQVSOEAKGTQGTGVEQTRLPATPNRSEP--H-----A 3329  
QY 2170 QRGAEPAEORNDAR-SPGSI SYLPSFTFKLENTSPMVKSKKQEIFRKLNSGSGGSDMAA 2228  
DB 3330 OLQRAVETAQAPAPSPVSVSMKFDLPSLSQA-----APKQPLFVPANS--GPSTPPGL 3383  
QY 2229 AOPGTEIFNLPAVITS 2244  
DB 3384 ALPHAEQVAPAKQESS 3399  
RESULT 12  
Q9NHNI  
ID Q9NHNI PRELIMINARY; PRT; 5554 AA.  
AC Q9NHNI;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE SPLIT ENDS LONG ISOFORM.  
GN SPEN OR CG18497.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kuang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;  
RT "Split ends encodes large nuclear proteins that regulate neuronal cell  
fate and axon extension in the Drosophila embryo.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF211715; AAF34661.1;  
DR FlyBase; FBgn0016977; spen.  
DR InterPro; IPR000504;  
SQ SEQUENCE 5554 AA; 599188 MW; 4037E27833D0C622 CRC64;

Query Match 3.8%; Score 484.5; DB 5; Length 5554;  
Best Local Similarity 18.1%; Pred. No. 3.7e-17;  
Matches 556; Conservative 389; Mismatches 1050; Indels 1085; Gaps 141;

QY 36 QEFAPVD--YRSSHLEVSQASQLLOQO---QQQLRRRPSLLSEFHPGSDRPOERR-----86  
DB 2583 QSQAPDLTIKQER-P1APAEIKREQLSDEQKFKSR-----HDSNSSIEERKLKTE 2634  
QY 87 -----TSYEPHPGP-----SPVDHDSLESKRPRLEQVSDSHFQVRYSAAVLPLV-130  
DB 2635 REIKTELGDFYNSSEYITTKLKEYSPETRKHKRKLKSSSTAD---TSAATPLVM 2691  
QY 131 HPLPEGL---RASADAKK--DPAFGGKHEAPSPISGQCGDDQNAQSPSKLSKEELIQSM 185



Db 2692 TPLTPIEDVHSSSECKTKFDNFDDLKTECCSIPLE--ISAGERRKHKKERKKKKLNNM 2750  
Qy 186 -----DRVDREIAKVEQOILKKKKOQLEEE-----AAKPPPE 220  
Db 2751 TEATVPNSPTINDTSSSEKLERHRLAKSKKSKMDNSCNTKIYNSSGAHPSTSPSLPA 2810  
Qy 221 KPVSPPV-----EOKHRISIVQIYDENRKKABEAHIFEGL-----GPKVE--L 263  
Db 2811 TPTSAPSTAQTSKRGEDKMEFIFGIISDEESQFPEQAETNKNDIIPSSVSTGTGIVSAAL 2870  
Qy 264 PLYNQPSDTK-----VYHENIKTNQVMRKKLILFFKRRNHARKQREOKICQRYD 312  
Db 2871 QTYKQEPSTPNKNEAAHIQTVHEPEOQQQLERLSSGSSSSSHADRERH----- 2922  
Qy 313 QLMFAWEKVDRIENNRPRKAKESKREYIEKQFEIRKQREQORFORVQORGAGLSAT 372  
Db 2923 -----RREKREKKRE-----KSQREQQNOI----- 2943  
Qy 373 IARSEHEISEIIDLSEQENNEKQMRQLSVIPPMFMDAEQRVKFINNGLMEDPMKYK 432  
Db 2944 -----HQSSKSVETKVDNDS-----VDMDEAGRALE-----AQLMSD----- 2976  
Qy 433 DRQFMNVWTDHEKEFEKDKETQHPKNFGLIASYLERKSVPCVLYLYLTKNENYKALVR 492  
Db 2977 -----FDTKPISEATPTAATY--RSDMTDVF--SDNE----- 3009  
Qy 493 RNYGKRRGRNOQIARPSQEEKEVEKEEDKAETKKEKEDEKEDKESKENTKEKD 552  
Db 3010 -----NNSVDMTKQGVSEQEQHKS--DKKKKKKSKKEKEKELLOQQRRESLP 3058  
Qy 553 KIDGTAETEEREQATPRKRTANSQGRKKGRITR----- 587  
Db 3059 NVASTSS-----APPTP--GLTVNVQAASKHADLQDAKHISPPVCKPSPSLPCLIGDD 3112  
Qy 588 -----SMTNEAAASAAAAATEEP--PPLPPP-----EPSTPEVTSRWTE 630  
Db 3113 DDDALHTPKAKPTTPSSNGDGLFPRKPKRLIPIPKPTIANSSTLSTQSAETPVSSG 3172  
Qy 631 BEMEYAKGLVHEGRNMAAIKRVG-----TKSEAQCK----- 663  
Db 3173 TVISSALATTPSTSTAAGVSAAPGLDNPSTPSASAQCKKESFIPGFDQLDDRISESAY 3232  
Qy 664 -NFYFNKRRINLNLQOHKQTSRKPREDVSCQESVAS--TVSAQDE----- 712  
Db 3233 QSISAENSTSLDNIADPRIPVAPSPRATKPLDKLEESKSRVTISOETESAVSALLG 3292  
Qy 713 -----DIEASNEE-----NPDESEVAVKPSEDSPENATSRGNT 748  
Db 3293 ESFGTSTTDSLDCMDEMSSVNELEPTLVIAEPDEEAALAAKAETAGEPASTL--EE 3350  
Qy 749 PAVELEPTTAPSTPSLAVPSTKPAP-----DESVEQVND--SISAETA----- 794  
Db 3351 P--EMEPERAEPPDPEAETESPPVVEVLDPPEELNKAQVSLKHEDMMDIKADTPQSERD 3408  
Qy 795 -QMDVDOEHSAGSCVCDPPATKASVDVVRVNPENHASKVEGDNTKRDLDRASEKV 853  
Db 3409 LQIDTDTEENPDE-----ADSSGSLKIDET-----VQSSSPKESISNNSP-- 3450  
Qy 854 EPRDELVVAQIINAQRPESQDNDSSATCSADEVDGEPERQRMFMDKPSLLNPTGS 913  
Db 3451 TPRETANIDPNVESQ--PKLSNESTP-----QPSVITKFLPDLTPKTV--PAG- 3495  
Qy 914 ILVSSPLKPNPLDLOLQRAAVIPPMVSCPTPCNIPITGPVSGYALYQRIKAMHESALL 973  
Db 3496 -LPPSPVKIEPTTSKLQ--PLVQPVQTVLPAPHSTGSGISANSVINLDSNVISSC-- 3550  
Qy 974 BEQRQREQIDLECRSTSCGTSKSPNREWEVLQPHAPHLITNLPEGVRLPTTRPRP- 1032  
Db 3551 -----SNTSASATASASASISFGSPTASQ--NAMPQASPPKQCPITPQ 3592  
Qy 1033 -----PPPLIPSSKTTVASEKPSFTMG--GSIQGTGCTYLTSHNQASYTQETPKP 1081  
Db 3593 QAIQTQSLMQPPTISIEQTPHFAVQMWLSPQSHHPQPGTYMVGIRAPSPHLSHP 3652

Qy 1082 S-----VGSIS--LGLPROQESAKSATLPYIKOEE-----FSPRSONSOPEGLLVRAQHE 1129  
Db 3653 GRGVAQSRVLQGLSPVGRPMVSPSPQOQVQTOQOHALITSPQSSNISP----- 3702  
Qy 1130 GVVRTAGAIQBSITRGTPTSIS-----VSIPSRLG--SITQGTTPALPORTGIPT 1180  
Db 3703 -LASPTTRVL-----SSNSPTTSKVNSYQPRNQVFPQSPKSAVEVQTPQL--MTIPLQ 3756  
Qy 1181 ALVKGSISRMPIEDSSPEKREAAKSHVIEGKSHGLLSYDNK--NAREGTRSPRTAH 1239  
Db 3757 KMTPIQVPHPTIILSVTVVQPOQATOSOVASSPPLGSLPHKVNHLNAHQNOQPOVTA 3816  
Qy 1240 EISLKRSYVESVGNKIQGMSMRSPVAPLEGILICRALPRGSPHSDKRTVLSSIMOG 1299  
Db 3817 KMTAQHQHQHQMFHQMQIQRQ-----QHMQOQQLHGOSOOITSAPOHQMHQ 3864  
Qy 1300 TPRATTESFEDGLKYPKO-----IKRESPIRAFEAGITKGPYDGIITTIKEMGRSI 1351  
Db 3865 QHQAOQOQHNNQOHLNQQHLAQOQHTOKHQAOQOQFNOQIQ-----QHOSQOQ 3913  
Qy 1352 HEIPRODILITQE-----SRKTPEVVQSTRPIIEGSIISQGTPIKFDN 1392  
Db 3914 HOVOQNOAQOQHLSSOQHQSOQQLNQOHAQOQLOQIKLQOHHGPOQOQKSPQGVGH 3973  
Qy 1393 NSGQSAI-----KHNKVSILITG-----PSKLSRGMPPLEIVPENIKVVERGYEDVKAGE 1442  
Db 3974 LGGSTSIASQOHNQOLPARGVPOOQHPOOLSHSP-----CKNTLVSVNQGVQPPAILTR 4030  
Qy 1443 TVRSRHTSVVSGPSVLSTLHEAPKAQLSPIGYDDTISARRTPVS--YQNT----- 1491  
Db 4031 VGSHSQPNQOQOLPHQOQSSSGHPHQQLSSPGA--NLQTPLVNIQNTPKIIVQOQHV 4087  
Qy 1492 -MSRGSPPMNRTSDVTIPPNKSTNHERKSTLTPTTORESIPAKS-----PV 1535  
Db 4088 AQNVQPPPTQOAGNAIHPQNOG-----KDSTPPGHVEPTPAMSAOKTSESIVIRTPPT 4142  
Qy 1536 PGVDVPSVSHSPDFPHRRGSTAGEV-----YWSHLPTQ 1567  
Db 4143 TGL-AVIS-----ANTVGSLLTEENLIKISQKODELIEQDSKESVDSDYNAKEVN 4192  
Qy 1568 LDPAMPFHRLDPAAYVLFORQL-----SPITGYPVSQOLYAMENRTOTI--L 1614  
Db 4193 IDSVI--KLDTPPLASKDAKRAVEMQAIAPAPENPQNGNSMAQETALPTTSMVNNS 4249  
Qy 1615 NDYITSQOMVNLRPDVARGLSPREOPLGLPYPAFPG-----IIDLTNMPPT 1661  
Db 4250 NDHDTEDTETRQLPPAKPPPTVTVGRPPGGSAGRGPRGAKKVGFPPLNSVTAAPPG 4309  
Qy 1662 I--LVPHPG--GTST-----PP-----MDRIT 1679  
Db 4310 VDSLWVQPDNGVQVTRLRKPVTAIVTVGRKGPPRPNLLQOQLOQOQDIDQKGMWMT 4369  
Qy 1680 YIPGTQITFPFPPPNYNSMSFGPHLAAAASAEEREREKERERERIAA--ASSDL 1736  
Db 4370 --SATSTPLPTPI-----PTSSVLTAAEKARNQALTAQOEONQVSAQVGTGDI 4418  
Qy 1737 Y--LRPGSEQPRGSGHYVRSPSPSVRTQE-----TMOQRPVS-----FGT 1778  
Db 4419 YEFHEDGGEEP--KAPT--ISSVAPSAEDQRPRLILTINKTOPSIKNISEMBQTTQOQOQ 4474  
Qy 1779 NGTSVITPLDPTA-----QLRIMPLPAG--GPSISQ 1807  
Db 4475 QQSEVISNTDPIGGDNSESCNTRKSRRLQEKEDRSTVDDIIVVNTNTPTGTGPHLPK 4534  
Qy 1808 GLPASRYNTAADALAALVDA-----AASAPQMDVSKTKESK--HEAARLEENLSRAA 1859  
Db 4535 GAQTPPRSRGRNAQAKTDAVQIINAVGRPR-----RSKDRKTIGEQTANLIEEVTASNAT 4590  
Qy 1860 V-----SEQOOLEQKTL----- 1872  
Db 4591 VAASHLAPPEGAGVESHVPQDLDAKEVEPVSVVVTPISTPAPVSAAPVTVPVAMVPVKPT 4650



```
QY 1873 -----VEKRSVOCLTSSAFPSGKPKQPHSSVVYS----- 1901
| | | | | : | | | | | :
Db 4651 MPQHPKKKAIAAEIESYQAI--NSSIPSGGLPMQTAAPAKITGKVADAVSKALVDP 4708
| | | | | : | | | | | :
QY 1902 -----EAGKDKG----- 1931
| | | | | : | | | | | :
Db 4709 VTGVTITAGMPQCKEGNLPAAATAAAPSNSNEDGAAPPQPPQHOQOQOHPQPPQOQANL 4768
| | | | | : | | | | | :
QY 1932 -FIDVITRIQIASDKDARERGSSQSSSLSSSHRYETPSDAIEVISPASSAPPQEKIQ 1990
| | | | | : | | | | | :
Db 4769 QINTWLISGLPNPITALGKSVQLETSAAALL-----NKPVSVLVKGNA-----SQVIQ 4817
| | | | | : | | | | | :
QY 1991 TVQPEVVKANOENDPTRO-----YEGP-----LHH-----YRPQQ-----ESPSPQOQ 2029
| | | | | : | | | | | :
Db 4818 QOQPOIVA-----PAKQPIILQONPLPTVLHHAQHTTVRPQPKAKAVLNREKNIQOQ 4870
| | | | | : | | | | | :
QY 2030 LPPSSQAEQMGVPTTHRLITLADHICQIITODFARNQ-----VSSQTPQO----- 2075
| | | | | : | | | | | :
Db 4871 LPPTQOAVAQ---PPQH-----APHSGHMLLTDTAGNQQLVQPIIARHLQOQOHLQVNV 4922
| | | | | : | | | | | :
QY 2076 -PPTSTFONSPALYSTPVRTKTSNRYSPESQASVHHORPGSRVS-----PENLVDKSR 2129
| | | | | : | | | | | :
Db 4923 PPTAHSPSPRI-----PSQOQO-----LGPASISPOQOQPTVVIKQA 4963
| | | | | : | | | | | :
QY 2130 GSRGKSPERSHVSSEYEPISPPQVYVHEKQDSLILLSQSG-----AEPAEQRNDARS 2184
| | | | | : | | | | | :
Db 4964 AS--AAQPOILHWSSKASVVPQO-----QQQLPPTSTSGPHLQAKP----- 5005
| | | | | : | | | | | :
QY 2185 PCSIYLSFFTKLENTSPWYKSKQOE-IFRKLNSSGGSDSM-----AAAPQTEI-F 2236
| | | | | : | | | | | :
Db 5006 --NYSYAP---TVLPPTLPFAVQOQOQOQHLKYKONNQKGAQIQMPHPHGLMLQO 5060
| | | | | : | | | | | :
QY 2237 NLPVATTSGSVSSRGHSTADPASNLGLEDIIRKALMGDFDKVEDHGVVMSQPMGVVPQT 2296
| | | | | : | | | | | :
Db 5061 KLPA-----HLQPOQHOL-NPSPPGKPNVL-----HGLSQO-----IMPGS 5098
| | | | | : | | | | | :
QY 2297 ANTSVVTSGTTRREGDSPSHGGVCK-----PKLISKNSRKS 2337
| | | | | : | | | | | :
Db 5099 V-----GSPPPVSAVLKTAQQQVNSVVPVAGIRTAIPNISQSPQSRVSPL 5144
| | | | | : | | | | | :
QY 2338 PIPGOGYLGTBRPSSVSVHSEGDY-----HRQ-----TP----- 2367
| | | | | : | | | | | :
Db 5145 VLP-PGISGV---PPFDASLHDLGAVSGRRQTQSPPPHQAQASPIPNBSTDYRGVTSRDF 5201
| | | | | : | | | | | :
QY 2368 -----GNAWEDRPSSTGSTQPPYPLNPLMRMLSSPTPTIACAPSAVNOAAPHQONR 2418
| | | | | : | | | | | :
Db 5202 MLYQHHLRGDGYDKMGSS-----PPELRRPGSGPPRTI-----AVPHS----- 5242
| | | | | : | | | | | :
QY 2419 IWREPAPLLSAQYETLSDS 2438
| | | | | : | | | | | :
Db 5243 -----LQSPQDRTAADS 5254
| | | | | : | | | | | :

RESULT 13
Q9VPL1
ID Q9VPL1 PRELIMINARY; PRT; 5560 AA.
AC Q9VPL1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SPEN PROTEIN.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
```

```
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bock J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003590; AAF51535.2;
DR Flybase; FBgn0016977; spen.
DR InterPro; IPR000504;
DR Pfam; PF00076; trm; 3;
DR Q9VPL1;
SQ SEQUENCE 5560 AA; 599996 MW; 58F19621AF40D2A8 CRC64;
```

## Query Match

3.8%; Score 484.5; DB 5; Length 5560;

Best Local Similarity 18.18; Pred. No. 3.7e-17;

Matches 556; Conservative 389; Mismatches 1050; Indels 1085; Gaps 141;

```
QY 36 QEFAPVD--YRSSHLEVSQASQLLQQQ---QQQLRRRRLSEFFHFGSDRPPQERR---- 86
| | | | | : | | | | | : | | | | | :
Db 2589 QSQPAPDLTIKQEH-PIAPAQEIKRQLSDEEQFKSR-----HDSNSEEKRLKATE 2640
| | | | | : | | | | | : | | | | | :
QY 87 -----TSYEPFHPGP-----SPVDHDSLESKRPRLEQVDSHFQORVSAVPLV- 130
| | | | | : | | | | | : | | | | | :
Db 2641 REIKTELGDYFNSEYTYTGKLEYSPETRRKKKKRRLKSSSTAD---TSAQTPLVM 2697
| | | | | : | | | | | : | | | | | :
QY 131 HPLPEGL---RASADAKK--DPAFGGKHEAPSPISGPGCGDQNASPSKLSKEELIQSM 185
| | | | | : | | | | | : | | | | | :
Db 2698 TPLPSPIDVHSSECKTKYFNDDLKTECSIPLE-ISAGERKKHKKERKKKKLKNM 2756
| | | | | : | | | | | : | | | | | :
QY 186 -----DRVDREIAKVEQOILKKKKQOOLEE-----AAKPPPE 220
| | | | | : | | | | | : | | | | | :
Db 2757 TEATVPNSPTTNDTSSEKLSKEERHLLKSKKSDMNSCNTKIYNSGSAHPSTSPSLPA 2816
| | | | | : | | | | | : | | | | | :
QY 221 KPVSPPPV-----EQKRSIVQIYDENKKKAAEAKHIFEGL-----GPKVE--L 263
| | | | | : | | | | | : | | | | | :
Db 2817 TPTSAPSTAQTSKRGEKMEFIFGIISDEESQPEAQETNKDIIPSVSTGPTVSAAL 2876
| | | | | : | | | | | : | | | | | :
QY 264 PLYNQPSDTK-----VYHENIKTNQVMRKKLILFFKRNHARKQEQKICQRYD 312
| | | | | : | | | | | : | | | | | :
Db 2877 QTYKQEPSTPNKNEEAHIQITVHEPQQOQLRSRSLSGSSSSSHADREH----- 2928
| | | | | : | | | | | : | | | | | :
QY 313 OLMDAEWEKKVDRIENPNRRKAKESKTREYEQPFEIRKQEQORFQVQRGAGLSAT 372
| | | | | : | | | | | : | | | | | :
```











Db 3502 -LPPSPVKIEPTISKLOQ---PLVQVQTVLPAPHSHTSGSISANSVINLDSLNVISSC--- 3556  
QY 974 EQRORQEQIDLECRSSYSPGCTSKSPNREWEVLQAPAHOLITNLPEGVRLTTPRTRP- 1032  
Db 3557 -----SNFSAASATASASISFGSPTASQ---NAMPQASTKQGPITPQ 3598  
QY 1033 -----PPPLIPSKTTVASRKPFSFIMG-GSISOGTPTGYLTSHNQASYTQETPKP 1081  
Db 3599 QAIRTQSLIMOPPTISPEQTFHFAVPQWLVSPQSHHPQCPQTYMVGIRAPSPHPLSP 3658  
QY 1082 S-----VGSIS-LGLPQROESAKSATLPYIKQEE---FSPRSONSOPEGLLVRAQHE 1129  
Db 3659 GRGVAQSLRGVLSVGRPMVSPQSPQOVQTOOQHALITSPQSSNISIP----- 3708  
QY 1130 GVRCTAGAIQSGITRGTPTSKIS-----VESIPSLRG-SITOGTPALPGTGPTE 1180  
Db 3709 -LASPTTVL-----SSNSPTTSKVNYSQPRNQOVQQSPKSAVEQITPQL-WTIPLO 3762  
QY 1181 ALVKGISIRMPEDSSPKGREAAKSHVYIEGKSHILSYDNK-NAREGTRSPRTAH 1239  
Db 3763 KMTPIQVPHPTIISKVTVQPOQATQSQVASSPPLGSLPPHKNVHLNAHQNOQOPQVIA 3822  
QY 1240 EISLARSYESVEGNKQGMRESVSPAPLEGLICRALPRGSPHDLKERTVLSGIMQG 1299  
Db 3823 KMTAHOHQHMQOFMHQMIQRO-----OHMQOQLHQGSOQITSAPOHQMHQ 3870  
QY 1300 TPRTATESFEDGLKYPKO-----IKRESPIRAFEGAITKGPYDGIITIKEMGRSI 1351  
Db 3871 QHOAQOQOHHNQHLNQLLHAQHPQKQHOAQOQFNQOIO-----QHOQOQO 3919  
QY 1352 HEIPRODILTOE-----SRKTPVQVQSTRPIIEGSIQGTPIKFDN 1392  
Db 3920 HOVOQOQAQOHLQSOQOQHOQOQOQLOQLOKLOMHGPOQOQKSPQGVGH 3979  
QY 1393 NSGOSAI-----KHNKSLITG-----PSKLSRGPPLLEIVPENIKVYRGKYEDVKAGE 1442  
Db 3980 LGGSTISIFASQOHSOLPARGVPOOHPQOQLSHSP-----CKPNTLVSNQGVQPPAILTR 4036  
QY 1443 TVRSRHTSVVSGSPVLSTLHEAPKAQLSPGIYDDTSARRTPVS-YQNT----- 1491  
Db 4037 VGSHSQPQOQOOLPHQOQSSGHGPHQKOLSSPCA---NLPLPTPLNVQNTPKLIIVQOQHV 4093  
QY 1492 -MSRGSPPMNTSDVTIPPNKSTNHERKSTLPTPTQRESIPAKS-----PV 1535  
Db 4094 AQONQVPPPTQGNATHYPQNOG-----KDSPPGHVEPTPAMSAOKTSBSVIRTPPT 4148  
QY 1536 PCQVDPVSHSPDPHHRGSTAGEV-----YWSHLPTQ 1567  
Db 4149 TGL-AVIS-----ANTVGSILTEENLIKISQKQDELIEQDSKEVDSYWSAKEYN 4198  
QY 1568 LDPAMPFFHRALDPAAAYLFQRL-----SPTPGYSPQVQLYAMENTROTII-L 1614  
Db 4199 IDSVI---KLLDTPPLASKDAKRAVEKQAIAPAPINPQGNQSMQOETALPTTSMVNS 4255  
QY 1615 NDYITSQOQVNLRPDVARGLSPREQPLGLPYPATRG-----IIDLTNMPPT 1661  
Db 4256 NDHDFEDETETROLPPAPPTPTVGRPPCGGSAKRGPRGAKKVGFPPLNSVTAAPPG 4315  
QY 1662 I--LYPHGC--GTST-----PP-----MDRIT 1679  
Db 4316 VDSLVOFGDNGVQTRLRKPVTPATVTRGKGRPPRNLLOQOQOQOQOQOQOQOQOQOQO 4375  
QY 1680 YIPGQITFPPRPYNASMSCPHPTHLAAASAEERERERERERERERERERERERERERER 1736  
Db 4376 --SATSSPTLPTPI-----PTSSVLTAEEKARNOALTQAOEQNOVASQVGTQGDII 4424  
QY 1737 Y--LRPGSEQPRGSHGVRSPPSVRTQE-----TMLQORPSV-----FOGT 1778  
Db 4425 YEFHEDGGECP-KPKT---ISSVAPSAEDQRPRLILITINKTPSIKNISEMEQTIQOQOQ 4480  
QY 1779 NCTSVITPLDPTA-----QLRIMPLPAG-GPISIQ 1807  
Db 4481 QOSEVISNTDPIGGDENSESCNTRKSRRLQEKEDRSTVDIIEDVVVRNTNTPTGTGPHLPK 4540

QY 1808 GLPASRYNTAADALAAALVDA-----AASAPQMDVSKTKESK---HEAARLEENLRSAA 1859  
Db 4541 GAOTTPRRSGRNAQAKKTDVQIINAVGRPR-----ASKRKTIQEGTANLIEEVTASNAT 4596  
QY 1860 V-----SEQOOLEQKTL----- 1872  
Db 4597 VAASHLAPPEGAGVESHVPOLDAKEVEPVSVVTPITPAPVSVAAADVTVPVAMVPKPT 4656  
QY 1873 -----VEKRSVQCLYTSSAFPSGKQPPHSSVVYS----- 1901  
Db 4657 MPQHPKKKAIAAEIESYQAI---NSSIPSGGLPMQHTAATAPATOKITGGVADAVSKALVDP 4714  
QY 1902 -----EAGKDKG-----PPKSYEEELRTGKTTITAA-- 1931  
Db 4715 VTGVTAGMPQCKEGNLPATAAPANSNEDQQAAPPQLOHQOQOQHPQPPQOQANL 4774  
QY 1932 -FIDVITRQIASDKDARERGSSQSSSSLSHRVETPDSDAIEVISPASSPAPPOEKIQ 1990  
Db 4775 QINTTLIPSGLPNIPITALGKSVQLETSAALL-----NKPVSVLVKGNA---SQVIQ 4823  
QY 1991 TYQPEVVKANOENPTRO---YECP---LHH-----YRPOQ-----ESPQOQ 2029  
Db 4824 QOQPOIVA-----PAKQPILOQNPLPTVLHHAQHTTVRPPQPKAHVLNREKNIOQO 4876  
QY 2030 LPPSQAQMGQVPRTHRLITLADHICQIITODFARNQ-----VSSOTPOQ----- 2075  
Db 4877 LPTTKQAVAQ---PQH-----APHSGHMLLTDAGNQLVQPIIARHLQOQOHLQVNV 4928  
QY 2076 -PPTTFQNSALYSTVPTVTKTSNRYSPESQAQSVHHORGPSRVS-----PENLVKSR 2129  
Db 4929 PPPTAHSHPRI-----PSOQOQ-----LPGASISPPQOQOQPVVVIKOA 4969  
QY 2130 GSRPKGKSPERSHVSSEPYEPIPPQVVPVHVEKODSLLLLSQSG-----AEPAEORDAKS 2184  
Db 4970 AS--AAQPOILHVSVKASVVPQO-----OQQLPPTSTSGPHLQAKP----- 5011  
QY 2185 PGSISYSPFFTKLENTSPMVKSKQE-IFRKLNSGGGSDM-----AAQPGTEI-F 2236  
Db 5012 --NYSYAP---IVLTPTLPVAVOQOQOQOHLKQNNQOKGAQIOMPHPHIIMPTHGMLQO 5066  
QY 2237 NLPAVTTSGVSRGHSFADPASNLGLEDIIRKALMGSDDKVEDHGVVMSQPMGVVPGT 2296  
Db 5067 KLPA-----HLOFQOQHL-NPSPPPGKPNVL-----HGLSQG---IMPGS 5104  
QY 2297 ANTSVVTSETRREGDPSPHSGVCK-----PKLISKNSRKS 2337  
Db 5105 V-----GSPPPVSAALVLTAAQOVNSVVPVAGIRTAIPNISPOSQPRVSP 5150  
QY 2338 PIPGQCYLGE-----RPSVSVSVHSEGDY---HROTPGNWEDRPSSTGSTQFPYNP 2387  
Db 5151 VLP-PGISGVPPFDASLNDSTYRGVTASRDFMLYQHHLMRGGDYDDKMGSS-----PP 5202  
QY 2388 LPMRLMSTPTTPIACAPSAVNOAAPHQONRIWEREPAPLLSAQVETLSDS 2438  
Db 5203 LELRRPGSGPPRTI-----AVPHS-----LQSPQDRTAADS 5233

## RESULT 15

Q9Y6V0

ID Q9Y6V0 PRELIMINARY; PRT: 3394 AA.

AC Q9Y6V0;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

DE WUGS:I\_DJ0784G16.1 PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99063792; PubMed=9847074;







Db 1534 S--APPVAP-----SSFOAAPTSTVQFLTTEVSKTEVSATRSAPSV--GL-----SSI 1579  
Qy 1319 KRESPIRAFEGAITKGPYDGIITIKEMGR-----SHEIPRODILTOESRKTPEV 1370  
Db 1580 SITIPPEPLADNIHLKP-----QYKEDGKLQLVGDVIDLRTVPKVEKTTD--KCIDL 1632  
Qy 1371 VOSTRPI---TEGSIQGTPIKFDNNSQSAIKHNKSLITGPKSLKSGMPLEIVPENI 1427  
Db 1633 SASTMDVKRQITANEVYKQI-----SAVQPSIIL--SVTSSIVTPVSLATETV 1680  
Qy 1428 KVERGKYEDYKAG--ETVRSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDISARTP 1485  
Db 1681 TFIYCTASASTTGTGESVLGAEHAMTT---PLQLTTSKHAEPYRIPSDQVFPFAAREAP 1737  
Qy 1486 VSYQNTMGRSP---MNMRTSDVTIPNKNSTNHERKSTL-----TPTQRE 1527  
Db 1738 IN---LSLGTPAHAVTAITKPTVPVPGVTNGTWDSTVSOGITDGEVVDLSTTKSHRT 1793  
Qy 1528 SIPAKSPVGV--DPVSHSPFDPHHRGSTAGE-----VYWSHLPTQLDOPAMPFHR-- 1576  
Db 1794 VVTMDESTSSVMTKIEDEKPDV-----LTAGRAVCCDVVY-----KLPFGRSC 1838  
Qy 1577 -ALDPAAAA-----YLFQR-----QLSPTPGYPSQYQLYAMENR 1610  
Db 1839 TAQOPATTLPEDRFGYRDDHYQDRSGPYGYRGICGMPKSPMSDNLAEAGHFFYKSKNA- 1897  
Qy 1611 QTLINDYITSQOMVNL---RPDVARGLSPREQLGLPLPATRGIIIDLTMNPPTILVPHP 1667  
Db 1898 ---FDYSEGTDTAVDLTSGRVTGEVMDYSKTTG--PYPETRQVIS-----G 1940  
Qy 1668 GGTSTPPMDRITYIGTOITPPRPYNSASMS-PGHPHTLAAAASAEERERERERER 1726  
Db 1941 AGISTP-----QYSTARMTPPPGPYCGVSV-----1966  
Qy 1727 ERIAASDLYLRGSEOPG-----RPGS--HGVRSPSPSVRTOETML-----QQ 1770  
Db 1967 --LRSSNGYSSVATPPTSTAITOPGSIFTSTVRDLS-GIHTADAVTSLPAMHSHOP 2023  
Qy 1771 RPSVFGTNGTS---VITPLDPTAQLIMPLPAGGSPISQGLPASRYNTAADALAAV- 1825  
Db 2024 MPRSFIITGASETDIATVIGDISASLTITMESLTATIDSVP--LTTASEVPEVVG 2081  
Qy 1826 --DAAASAPQ-----MDVSKTKE-----SKHEAARL-----1849  
Db 2082 DESALLIIVEEDKQOQDLDERELLEKEIKOORFAEELEWEROEIQRFEGEKTWQVK 2141  
Qy 1850 -----ENLRSRSAVSEQOOLEKT-LEVEKRSVQCLY 1882  
Db 2142 LEEQSMKQHLLFOQEEERQAQFMROETLAQOQLQLEQIQOQLQHQLEEQKIRQIY 2201  
Qy 1883 TSSAPFPGCKPOPH-----SSVYSEAGKDKG-----1908  
Db 2202 QYNVDPSTASPTQTEQAILEGQYAALEGQFQWATEDATTTASAVVAIEIPQSGWTV 2261  
Qy 1909 -----PP-----PKSRYEEELTRGKTTITAAANFID---1934  
Db 2262 QSDGVTVQIAPPGILLSTVSEIPLTDVVVKEEKQPKR--SSGAKRGVQYDDMGENTDDPR 2320  
Qy 1935 ---VITRQIASDKDAREG--SQSSDSSSLSHRYETSDALEVISPASSAPPOEK 1988  
Db 2321 SFKKIVDSGVQTDDEATDRSYVRRRTTKKSVDTSVQTDDEDDODWDMPTRSR--KAR 2378  
Qy 1989 LOTQOPEVVKANOENDPTROVEGLHHYRPOQESPPQOQLPPSSQAGMGQVPRTHRL 2048  
Db 2379 VKYGDSTMEADKYK-----PL-----SKVSSIAVQTVAEISVQTEPVGTI--RTESI 2424  
Qy 2049 ITLADHICQITQDFAFN-----QVSSQTPQOQPTSTFQNSALYSTVPVTKTS 2098  
Db 2425 RARVDKAVEIKHISAPEKTYKGSGLGCQTEADSDTQSPQVLSATSPPKDKRP--TPLE 2482  
Qy 2099 NRYSPESAQSVHHQRGSRVSPENLVDKSRGSRPGKSPERSHVSSEYE-----PISP- 2152  
Db 2483 IGYSSHLRADSTVOLAPSPKPVLYSPISPLSPGSALESAFV---PYEKPLPDDISPO 2539

Qy 2153 -----PQVPVYHEKQDLLLLLSQGAEPABQRNDARSPGISYLPSPFTTKLENTSPMV 2205  
Db 2540 KVLHPDMAKVPPASPPTAKMQORSMDPKPLSPTADESSRAPFQYEGYTTKGSQTM--- 2596  
Qy 2206 KSKQOEIRKLNLSNGGSDMAAOP-----GTEIFNLPAVTTSGSVSSRGHSPAD 2256  
Db 2597 -----TSSGAQKVKRKLTPNPPPEISTGTQ-----STFTMGTVSRRICRTN 2640  
Qy 2257 PASNLG--LEDIIRKALMGSDDKVEDHGVMQSPGVVPG-----TANTSVVTSGETRE 2310  
Db 2641 TWARAKILQDIDREL-----DLVERSAKLKKQAELDEEKEIDAKLYLEMGINRRK 2694  
Qy 2311 EGDPSPHSGGVCKPKLISKNSRKSPIPG-----QGYLGTERPSSVSSVHSEGD 2361  
Db 2695 EA-----LLKEREKREAVLQGVAEQDRDYMSDSEVSSTRPTRIESQHG--- 2737  
Qy 2362 YHRQTPGWAWEDRPSSTGSTQF-----PYNPLTMRMLSSSTPTPIACAP 2405  
Db 2738 -----IERPTAPQTEFSQFIPPTQTESQLVPPTSPTQYQYSS-PALPTQ-AP 2785  
Qy 2406 SAVNOAAPHOQNRIRWEREPAPLLSAQYET 2434  
Db 2786 TSYTQOSHFEQOTLHQOVSP-----YOT 2809

## RESULT 16

Q9UQ35 PRELIMINARY; PRT: 2752 AA.

AC Q9UQ35;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE RNA BINDING PROTEIN  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeiki K., Sawada Y.;  
RT "Homo sapiens mRNA for RNA binding protein, complete cds."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016092; BAA83718.1;  
DR InterPro: IPR002965;  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 2752 AA; 109C64F181097123 CRC64;

## Query Match 3.7%; Score 473; DB 4; Length 2752;

Best Local Similarity 19.2%; Pred. No. 6e-17;

Matches 554; Conservative 316; Mismatches 1121; Indels 890; Gaps 122;

Qy 33 RHQGEFAVDYRSHLEVSAQSLQOQQOQOQLR-----RRPILLSSEPHGS 79  
Db 247 RSRSTTPAPKSRRAHRSATSADSSSDTSRSRSRAAAKHTTALAGRSPS-----PAS 300  
Qy 80 DRPQERRTSYEPF-HPGSPVDHDS---LESKRPRLEQVSDSHFQVSAAVLPLVHLPE 135  
Db 301 GR---RGEGDAPFSPGTTQRPSPSPETATKQSPSPVEDDKDKKKSATRP--SPSPE 355  
Qy 136 GLRASADAKKQPA-----FGGKHEAPSPISGQPCGDDQDNASPSK-----LSKEEL 181  
Db 356 ---RSTGPEPPAPTPLLAERHGGSPQPLATPTLSQEPVNPPEASPTDRSPKSPKEL 412  
Qy 182 IQSMDVRDREIAKVEQOILKLLKKQOOLEEAAKPEPEKPVSPVPEOKHRSIVQIYD 241  
Db 413 PQSS---SSSESPSPQPTKVSr-----HASSSPESPCKPA---PAPGSHREISSPTS 459  
Qy 242 ENRK---KAEAAHKIFELGLPKVELPLYNQPSDVKVHENIKTNQVMRKLLILFFKRN 297  
Db 460 KNRSHGRAKRDKSH-----SHTPSRRMGRSRSPATAKGRSRSTP-----TKRG 504















Qy	235	----	IVQIYIB-	----	NRKKAEEAHKTFEGLGPKVELPLYNQ	269					
		:	:	:	:	:					
Db	1361	PSETRIDIS	EEIEKEISQEKVTSKKD	SAQGPSPSRKEHENPELV	DDLSR-	1411					
		:	:	:	:	:					
Qy	270	SDTKVYHENIKY	NVMRKKLILFFKRR-	----	NHARKOREQIKORYQOLMEAMEKKVDRI	325					
		:	:	:	:	:					
Db	1412	----	RASYDS	VESSSENSPVARRRRKTSIGSSSE	EYKDSOGSGEDEFTRKQI--	1466					
		:	:	:	:	:					
Qy	326	ENNPRAKESK	TREYIEKQFPEIRKQEQERFOR-	----	VGORGAGLSATIARS	376					
		:	:	:	:	:					
Db	1467	EMSADEDA	SGSDEEFIRSQ	KEIGVTVESQKRETKGKSPAGKHRRLLTRKS	SSTSFDD	1526					
		:	:	:	:	:					
Qy	377	----	PHETSEIIDGLSE-	----	OENNEKQMRQLSVIPPMFMDAQRKRVKFINNMGLME	425					
		:	:	:	:	:					
Db	1527	DAGRRHSW	DEDETFDESPLKFR	ETKSQSESELVAG-	----	GGGLRRRKTIELNSTVT	1582				
		:	:	:	:	:					
Qy	426	DPMKVYKDRQ	WMVWTDHBEKFKDKFIQHPK-	----	NFGLIASYLERKSPVDCVLIY	478					
		:	:	:	:	:					
Db	1583	DKYSABSSQ	KTWLTFDEPELEMESL	TDSPEDRSGEGSSSLHASFTTGT	SPTS--S	1640					
		:	:	:	:	:					
Qy	479	YLTKNE-	----	NYKALVRRNYGRRRNOQIARPSOE	KEVEKEDEKAEKTEKKEEKKDE	535					
		:	:	:	:	:					
Db	1641	SLDEDS	SDSPSHKGESQOKARHS	HGPLLPTIEDSSSEBELREBEL	LKQEQOREL	1700					
		:	:	:	:	:					
Qy	536	BEKDEKED	SKENTKDKIDGTAETEBREQATPRGRKTANSQ	QRRKGRITRSMWNEAAA	595						
		:	:	:	:	:					
Db	1701	EQOQRSSSK	KKDKDEL--	RAQRRRRPKTPP-	----	SNLSPIEDASPTTEL-RQAAE	1752				
		:	:	:	:	:					
Qy	596	ASAAAAATE	PPPLPPPE--	PLSTE-	----	PVETSRW--TEEMEVAKKG-	639				
		:	:	:	:	:					
Db	1753	MEELHSSC	SEYSPSIEDPGEFELS	PEKIEVOKVYKLP	TAVSLYSPTEQSQVMQKEA	1812					
		:	:	:	:	:					
Qy	640	----	----	----	----	-----LVEHGRNW	647				
		:	:	:	:	:					
Db	1813	QKALKSAE	WYBEMMHKPHYKAP	PAANERDEVEKEP	LYGGMLEIDYIYESLVEDTYNG	1872					
		:	:	:	:	:					
Qy	648	AATAKMV	GKTSBAQCKNFYNYKR	RHNDLLOQ-	----	HKQKTSRKPREEDRVSQCESV-	702				
		:	:	:	:	:					
Db	1873	SVDSGL	LTQRDE--	QNGFMQORGREQIRLREQIYDDPMOKITD	LQKEFVELESLSIV	1929					
		:	:	:	:	:					
Qy	703	----	----	ASTVSAQDE-	----	DIEASNEE-	720				
		:	:	:	:	:					
Db	1930	QOEDIVSS	YIIPESHEIVDLGSMV	TSPSEKLLDDAAYEELMKQ	QMVGDSGLIQ	1989					
		:	:	:	:	:					
Qy	721	----	----	ENPEDSEV	AEVAKPSDESPENATRGNTE-PAVEL-	-----	753				
		:	:	:	:	:					
Db	1990	TTMGDDMA	BESTLDFDRVQDAS	LTSSILSGASLTOSTSATLSIPDVKITOH	FSTEEFEDE	2049					
		:	:	:	:	:					
Qy	754	----	----	EPTTTAP	STGPSLAVP--STKPAEDES	VETQVND	786				
		:	:	:	:	:					
Db	2050	YVTDYTR	IEQIIAHESLILATYSEPS	SATSVPPSDTPSLTSSISSVCTT	DDSSSPVTLD	2109					
		:	:	:	:	:					
Qy	787	SISAETA	QOMVDYQOEHSAE-	----	GSVCDDPP-	-----	815				
		:	:	:	:	:					
Db	2110	SLTVTTE	PADVIATFKDSE	ISSYFPFGSVIDYPEDIGVSLDR	TIPEISRTNADQIMIS	2169					
		:	:	:	:	:					
Qy	816	----	----	ATK-	----	ADSDVDEVVRYPENH-	----	ASKVEGNTKEROLDRASEKVEP	855		
		:	:	:	:	:	:	:	:		
Db	2170	FPCIAP	SITESVATPER	POADTISTDLPISEKELIKGKETG	DCIILEVLDAYK	KKREE	2229				
		:	:	:	:	:	:	:	:		
Qy	856	RDEBLV-	----	VAQOII	NAORPEQSDNDSSATCSADEDVGE	PERQMPDMSKPSLLNPTG	912				
		:	:	:	:	:	:	:	:		
Db	2230	SEAE	LTKISL	PETGLATPSSQTK	BOQPSHVS	GSEISGQ-	-----	EKPYRS	PSG	2279	
		:	:	:	:	:	:	:	:	:	
Qy	913	SILVSS-	PLKPNP-	----	LDL-	-----	----	PQLQ	HRAAVIP	PWMSCTPCNIP	949
		:	:	:	:	:	:	:	:	:	:
Db	2280	GLPV	STHPSKSHFF	RRSSSLDISA	QPPPPPPPPPPPPPPPPPPPPPP	PPPPPPPPPPPPPPPPPPPPPP	PPPPPPPPPPPPPPPPPPPPPP	PPPPPPPPPPPPPPPPPPPPPP	PPPPPPPPPPPPPPPPPPPPPP	PPPPPPPPPPPPPPPPPPPPPP	2339
		:	:	:	:	:	:	:	:	:	:
Qy	950	----	IGTPV	SGYALYQRH	IKAHESALLEBQRQ	REQIDLECRSS	TSPCGTSK	SPNREW	1004		
		:	:	:	:	:	:	:	:	:	:
Db	2340	KRKLAA	APVAPTIVTA	HADAIP	TVETAARRSNG	-----	----	LPAT	KICAAAPP-	-----	

Qy	1005	EVLOPAPHQLITWLP	PGVGLP	TRTRPP	PPPLIP	-----	SSKTTVAEKS	PFIMG	SGISQ	1050												
Db	2389	--VPPKPS	SIPTGLV	TFH	PEAKSP	PIAPKPA	VP	PIPTV	TQKTTD	CPKPT	2437											
Qy	1060	GTCTYLTSHN	--QASY	QTEY	PKPSV	SGSIS	LGL	PRO	ESAKS	ATLYIKOE	ESP	1116										
Db	2438	GLP	UTSNMS	UNLV	TSADY	--KL	PSPT	-----	SPL	SPHSNKS	-----	SPRYK	2477									
Qy	1117	SOPEGLIV	--RAQ	HEG	WVRG	TAGAI	QEGS	ITRG	TPSK	-IS	VE	SIP	LSR	GSI	TG	OT	GP	ALP	1173			
Db	2478	SLMET	VVIT	LP	SEP	GT	PTD	SSA	QA	LTS	WPL	SG	SPK	DLV	SL	VE	FSV	-----	VP	PM	2530	
Qy	1174	QTGIPTA	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1207	
Db	2531	STEIP	SASO	PTLY	TSG	ALG	TF	SV	TPAV	TAS	LFT	Q	VL	PA	EAK	S	VE	SA	VS	AVP	2590	
Qy	1208	GHVIE	KSG	HILL	SYD	NI	KNA	REG	TRS	PTA	HE	IS	LK	S	VE	SGN	I	KQ	MS	RES	1267	
Db	2591	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2619	
Qy	1268	PLEG	--LIC	RAL	PRG	SPH	DLK	--ERT	-L	SGS	I	MOG	T	PRAT	T	ES	F	ED	GL	KY	1322	
Db	2620	PLIG	DA	LD	LR	TPK	-----	SEV	KTE	K	C	D	L	S	A	M	V	K	R	O	2670	
Qy	1323	PPIRA	E	G	A	I	T	K	G	P	V	D	G	I	T	T	K	E	M	R	S	1382
Db	2671	PSI	IN	LSA	S	A	S	I	L	G	T	-----	-----	-----	-----	-----	-----	-----	-----	-----	2699	
Qy	1383	SOCT	PIK	F	D	N	S	Q	S	A	I	K	H	N	K	V	S	L	T	G	1442	
Db	2700	--DT	TY	T	G	T	ES	O	G	I	E	H	A	V	T	S	-----	-----	-----	-----	2724	
Qy	1443	TV	SR	HT	S	V	V	S	G	P	S	V	L	R	T	L	H	E	A	P	1502	
Db	2725	-T	YS	K	H	T	E	L	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2767		
Qy	1503	SD	V	T	I	P	N	K	S	T	N	H	E	R	K	S	T	I	T	P	1545	
Db	2768	K	P	V	T	P	V	G	V	T	G	W	T	D	S	I	O	G	I	T	2827	
Qy	1546	P	F	D	H	R	G	S	T	A	G	E	-----	-----	-----	-----	-----	-----	-----	-----	1584	
Db	2828	P	V	D	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2872		
Qy	1585	Y	L	F	Q	R	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1627		
Db	2873	Q	Y	D	R	S	G	P	Y	R	G	I	G	M	K	S	M	S	D	N	2927	
Qy	1628	R	P	V	A	R	G	L	S	P	R	E	O	G	L	P	V	A	T	R	1687	
Db	2928	R	V	S	T	G	E	M	D	Y	S	S	K	T	G	-P	P	E	T	R	2968	
Qy	1688	F	P	R	P	-Y	N	S	A	S	M	S	G	P	H	T	A	A	A	S	1746	
Db	2969	P	P	G	P	Q	G	V	S	V	-----	-----	-----	-----	-----	-----	-----	-----	-----	2998		
Qy	1747	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1787		
Db	2999	T	F	A	I	T	Q	P	G	S	I	F	T	V	R	D	L	S	G	H	3058	
Qy	1788	D	P	T	A	L	R	I	M	P	L	P	A	G	P	S	I	S	Q	G	1831	
Db	3059	D	I	N	A	S	L	O	T	I	T	-----	-----									



Db 3231 TEAILEGQYVATESQFATEDATTTASTVAIEI-----PQSQGWYTVQSDGVYQY 3283  
Qy 1926 -----TIAANFIDVILITROASDARERGSSDSSSLSSHHRYETPSDAIEVIS 1977  
Db 3284 IAPPGILSTVSEIPLDVTVVKEEKPKK--RSGGAKRGQYDEMGESMADPRNLKLIYD 3341  
Qy 1978 PASSPAPPOELQTY-----QPEVVKANQANDPTROYEGPLHHVRPQO-----ESP 2024  
Db 3342 SGVQTDDEETADRIYASRRRTKSVDFSVQTDDEQDEWMPRSRRKARTGYGDSGA 3401  
Qy 2025 SPQOOLPPSS-----QAEGMGQVPRTHRLITLADHICQII-----TODFA 2064  
Db 3402 EGDTKPKPSKVSSVAVQVVAEISVQTEPLGTI-RTPSIRARVDAKVEIKHISAPEKTYK 3460  
Qy 2065 RNOVSSOPPOPTSTFONSPSALVSTPVRTKTSNRYSPESQAQSVHHQR-----PG 2116  
Db 3461 GSGLCQETETPDPT-----QSPPYMGATSPKDK--KRPTPLEIGYSSSHLRADPTVQLAPS 3515  
Qy 2117 SRVSPENLVKSRGSRPGKSPERSHVSSEPYE-----PISP-----POVPVVHEKQD 2163  
Db 3516 PKSPKVLVSPISPLSPGHLEPAFV---PYEKPLPDDISFQKVLHPDMAKVPASPKTA 3572  
Qy 2164 SULLLSQGAEPABQRNDARSPGISYLPSPFTTK--LENTSPMVKSKKQEIIPKLNSSGGG 2222  
Db 3573 KMQORSMDPKPLSPADESSRAPFOYSEGFTAKGSQTTSGTKQKVRKL-----P 3623  
Qy 2223 DSDMAAQPTEIFNLPAVTTTSGSVSSRGHGFADPASNLG--LEDIIRKALMGSDDKVED 2281  
Db 3624 NPPPEASTGTQ-----STYSTMGASRRMCRNTMTARAKTLQDIDREL-----DLVER 3673  
Qy 2282 HGVNMQPMGVVPG-----TANTSVTSGETRRREGDPSPHSGGVCKPKLISKNSRKS 2336  
Db 3674 ESALRKQAEDEEKEIDAKLYLENGINRKEA-----LLKREKRER 3719  
Qy 2337 SPIPG-----QGYLGTERPSVSVHSEGDYHROTPOGAWEDRPSSTGTQF----- 2383  
Db 3720 AYLOQVAERDYMDSSESVSTRPSVESOHG-----IERPTAPQTEFSOFI 3766  
Qy 2384 -----PYNPLTMRMLSTPTPIACAPSAVNQAAPHQNRNWEREPAPLLSA 2431  
Db 3767 PPQQTQEAQLVPTSPYQYQYSS--PALPTQ-APTPTQOQHFOOTLYHOOVSP----- 3819  
Qy 2432 YET 2434  
Db 3820 YQT 3822

## RESULT 19

Q9QYX7 ID Q9QYX7 PRELIMINARY; PRT; 5038 AA.  
AC Q9QYX7  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE ACZONIN.  
GN ACZ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99439764; PubMed=10508862;  
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
RA Kilimann M.W.;  
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
RT zones, shares homology regions with rim and bassoon and binds  
RT profilin.";  
RL J. Cell Biol. 147:151-162(1999).  
RL EMBL; Y19185; CAB60731.1; -.  
DR HSSP; P04410; 1A25.

DR MGD; MGI:1349390; Acz.  
DR InterPro; IPR000008; -.  
DR InterPro; IPR001427; -.  
DR InterPro; IPR001478; -.  
DR Pfam; PF001168; C2; 2.  
DR Pfam; PF00595; PD2; 1.  
DR ProDom; PD000535; -; 1.  
DR ProSITE; PS00499; C2\_DOMAIN\_1; UNKNOWN\_1.  
DR ProSITE; PS00004; C2\_DOMAIN\_2; -.  
DR SMART; SM00239; C2; 1.  
SQ SEQUENCE 5038 AA; 547541 MW; 577A7B2530F54C5C CRC64;

Query Match 3.7%; Score 467.5; DB 11; Length 5038;  
Best Local Similarity 18.3%; Pred. No. 2.7e-16;  
Matches 562; Conservative 393; Mismatches 1079; Indels 1029; Gaps 133;

Qy 21 PPHSVQVTFNTRHQQFAVP-----DYRSHLEVSAQSL-----LQOQOQOQLR 66  
Db 1140 PLEEKPIPDQKLPDPDAKPSASEGEKRDLLKAHVQIPEEGPIGKVASLACEGQOQPD 1199  
Qy 67 RPSLLSEFHPGS-----DRPQERTSYEPFHPGSPVDHDSLEKRPRLQVSDSHF 119  
Db 1200 TRPEDL-----PGATPQTLPKDRQKESRDVTPQQAEGTAKEGRG--EPSKDRTEEKDK-- 1251  
Qy 120 QRVSAAVPLPLVHPLPEGLRASADAKDPAGGKHEAPSSPSISGQPCGDDQNASPSKLSKE 179  
Db 1252 SDTSSSQP---KSPQGL-SDTGYSSDGISSGLGEISLPS-----DEKDLKGLKKD 1301  
Qy 180 ELIQ-SMDVRDREIAKVEQQLKLKKKQQ--LEEEAAKPEPEKVPSP--PPVEQKHRS 234  
Db 1302 SFSQESSPSSDLAKLESTVLSILEQAASLTGVEKAEEKTPQK--VSPEQPQDQKQTOT 1360  
Qy 235 -----IVQIYDE-----NRKAAEAHKIFEGGLGPKVELPLNQ 269  
Db 1361 PSEPRDISIEEIKESQEKVTSKDSAQGFPSRKEHKNELVDLSPR----- 1411  
Qy 270 SDTKVYHENTKTQNMVRKKLILFFKRR-----NHARKQEKQICQRYDQLMQMEAEKKVDRI 325  
Db 1412 ---RASYSVDSSESSENSPVARKKRTSIGSSSEYKQEDSOGSEDEDFIRKQI--I 1466  
Qy 326 ENNPRKAKESKTRYEYKOPPEIKRQREOQERFOR--VGORGAGLSATARS----- 376  
Db 1467 EMSADEDASGSEDEEFIRSQLKEIGVTESQKREETKKGKSPAGKHRLTRKSTSFDD 1526  
Qy 377 -----EHEISELIIDLSE---QENNEKQMRQLSVPPMMFADAEQRRYKFINNGLME 425  
Db 1527 DAGRRHSHWDEDETFDESPELKFRTKSOESELVWAG-----GGGLRRFKTIELNSTVT 1582  
Qy 426 DPMKVYKDRQFMNVWTDHEKEIFKDKFIQHPK-----NFGLIASYLERKSVDPCLVLY 478  
Db 1583 DKYSAESSQKKTLYFDPEPELEMESLTDSPDRSGEGSSSLHASFTPTGTSPTSV--S 1640  
Qy 479 YLTKKNE---NYKALVRNRYGKRRNQQTARPSQOEKVEKEDEKAKTEKKEEEKKDE 535  
Db 1641 SLDEDSDSPSHKKGESQKQKQARRSHGHPLLPTIEDSSESEELREBELLEKQEKQEL 1700  
Qy 536 EEKDEKEDSKENTKEKDKIDGTAETEEREQATPRGRKTSANSGRRKGRTRSRMTNEAAA 595  
Db 1701 EQQQRKSSSKSKKDKDEL--RAQRRRERKPTTP-----SNLSPIEDASPTTEL-RQAAE 1752  
Qy 596 ASAAAAATEEPPLPPPPPE--PISTE-----PVTSRW--TEEMEVAKKG- 639  
Db 1753 MEELHRSSCSEYSPSIESDPGEPFISPEKIIQVKVYKLPATVSLYSPTDEQSVMQKEGA 1812  
Qy 640 -----LYEHGRNW 647  
Db 1813 QKALKSAEMEYEMMHKPHYKAPPAANERDEVEFEKEPLYGGMILIEDIYIESLVEDY 1872  
Qy 648 AATAKWGTGKSEAQCKNFYFNKRRHLDNLQO-----HKQKTRKPREERDVSCQESV- 702  
Db 1873 SVDSGLTRQDE---QNGFMPQREQKIRLREQIYDDPMQKITDLQKEFEYESLSHIV 1929



QY 703 -----ASTVSAQED-----DIEASNEE----- 720  
Db 1930 POEDIVSSYIIPESHEIVDLGSMVTSTSEBKLLDADAAYEELMKRQOMQVTDGSSLIQ 1989  
QY 721 -----ENEDSEVEAVKSEDSPENATSRGNTPE-PAVEL----- 753  
Db 1990 TTMGDDMAESTLDFRDVODASLTSSILSGASLTOSTSATLSIPDVKITQHFSTEPEFEDE 2049  
QY 754 -----EPTTETAPSTPSLAAP-STKPAEDSESVETQND 786  
Db 2050 YVTDYTRIEQEIHAHESILITYSEPSSESATVPSPDPSLTSSISSVCTTSSSPVTLLD 2109  
QY 787 SISAEATQMDVDQOEHSAE-----GSVCDPPP----- 815  
Db 2110 SLATVYTPADVITKFKDSESISSITYPGVSIDYDIPEDIGVSLDRITIPESRTNADQIMIS 2169  
QY 816 -----ATK-----ADSDVDEVVRYPENH-----ASKVEGDNTKREDLDRASEKVEP 855  
Db 2170 FPGAPSITESVATKPERPOADTISTDLPISEKELIKKGTGDIILEVLDAYKDRREE 2229  
QY 856 RDEDLV---VAQOINAOPEQSDNDSSATCSADEVDGEPERQRMFPMDSKPSILLNPTG 912  
Db 2230 SEAEITKISLPETGLATPPSQTKEQPGSHSVSGEISGO-----EKPTYRSPG 2279  
QY 913 SILVSS-PLKPNP-----LDL-----POLQHRAAVIPPMVSTPCNIP 949  
Db 2280 GLPVSTHPSKSHPPFRSSSLDISAQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 2339  
QY 950 -----IGTPVSGYALYORHAKHESALLEEORQORQOQIDLECRSSTSPCGTSKSPNREW 1004  
Db 2340 KRKLAAAPVAPTAVTAHADAIPTVBATAARRSNG-----LPATKICAAAPPP----- 2388  
QY 1005 EVLOPAPQLITNPEGVRLPTTPTPTPPPPPLIP-----SKTTVASKEKPSFIMGGSISQ 1059  
Db 2389 --VPPKSSIPTGLVTHREPEASRPPIAPPAVPEIPVTQKTTDTCPKPI----- 2437  
QY 1060 GTPQTYITSHN---QASYQTQTPKPSVGSISLGLPROQESAKSATLPYIKOEESPSRSON 1116  
Db 2438 GLPLTNSMLNLTADY---KLPSPT-----SPLSPHSNKS-----SPRYSK 2477  
QY 1117 SQPEGLLV--RAQHEGVVRGTAGAIQESITRGTPTSK-IVESIPSLRGSIOTGPALP 1173  
Db 2478 SLMETYVITLPSPEGPTDSSAAQAITSLWPLGSPPKDLVLETFSV-----VPPMT 2530  
QY 1174 QTGIPTEA-----LVKGSISR-MPIEDSSPEKGREEAASK 1207  
Db 2531 STEIPSASQPLTYTSGALGTFSVTPAVTASLFQVPTSLTQFLPAEASKPEVSASVAVP 2590  
QY 1208 GHVIEKSGHILSYDNKINAREGTRSPRAHEISLKRSESVESGNIKQGMRESVPVA 1267  
Db 2591 -----SVAPRSVSIPPEPLALDRHQYKENGKL----- 2619  
QY 1268 PLEG--LICRALPRGSPHDLK--ERTV-LSGSTMOGTPTATTESFEDGLKYPKQIKRES 1322  
Db 2620 PLIGDAIDLRTIPK-----SEVVKTEKMDLSASAMVYKQRTANEV-----YRROI SAVQ 2670  
QY 1323 PPIRAFEGAITKGPYDGIITIKEMGRSIIHEIPRODILTQESRKTPEVVOQSTREPIEGSI 1382  
Db 2671 PSILNLSAASLGTP-----VTMDS-KTAVAVTCT----- 2699  
QY 1383 SQGTPIKFDNNSGQSAIKHNKVSILITGPSKLSRGMPPLEIVPENIKVVERGKYEDVRAGE 1442  
Db 2700 --DTTIVTTGTSQVGHIEHATS-----PLQL----- 2724  
QY 1443 TVRSRHTSVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYONTMSRGSPMNRPT 1502  
Db 2725 --TTSKHTEL-----QYRKPSQAFPMIRDE-----APINLSLGPSTQAVTLAVT 2767  
QY 1503 SDVTIPPKNKSTNHERKSTLTP--TQRESI---PAKS--PVPBGVDVPSH-----S 1545  
Db 2768 KPVTPVPVGTNGWTDSTISQGITDGEVDLSTSKSHRTVVTWMDSTSNVVTKIIEDEEK 2827  
QY 1546 PFDPHRGSTAGE-----VYWSHLPTQLDPAMPFHR---ALDPAAAA----- 1584

Db 2828 PVD-----LTAGRRAVCCDMVY-----KLPFGSRCTAQOQAPATTLPEDRFGYRDDH 2872  
QY 1585 YLFOR-----QLSPTPGYPQYOLYFAMENRTQILNDYITSQOMQVNL----- 1627  
Db 2873 YOYDRSGPYGYRGYGGMKPSMSDNLAEAGHFFYKSKNA-----FDYSGGTEAAVADLTSG 2927  
QY 1628 RPDVARGLSPREQPLGLPYPATRGIIDLNNPPTILVPHPGGTSTPPMDRITYIPGTQIT 1687  
Db 2928 RVSTGEVMDYSSKTTG-PYPETROVISGV-----GISTPQYST-----ARMT 2968  
QY 1688 FPPRP--YNSASMSPGHPTHLAAASAEERERERERERERERERERERERERERERERER 1746  
Db 2969 PPPGQYGVGSV-----LSSNGVWYSSVATPIPS 2998  
QY 1747 -----RGS--HGYVRSPS-----PSVTOETMLQORPS-----VFOGTNGTSV-ITPL 1787  
Db 2999 TFAITTPGSIPTSTVTRDLSGIHTDAITSLALHQSPRPSYFITTGASSETDISVTSI 3058  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYN-----TAADALAALV-----DAAASA 1831  
Db 3059 DINASLOTITM-----ETLPAETMDSVPTLTASEVFSEVVEESTLLIJPDEDKQ 3110  
QY 1832 PQMDVSK-----TK 1840  
Db 3111 QOLDLERELLELEKIKOORFAEELEWEROETORFREOEKIMVQKKLEELQSMKQHLLYQ 3170  
QY 1841 ESKHEAARL--EENLRSRAVSEQQOLEOKT-LEVEKRSVOCLYTSSAFSGRQPH-- 1895  
Db 3171 EERQAOFMKROETLAQQQLQLEQIQOOLQHOEQKLRQIQYVNYEPSTASPOTT 3230  
QY 1896 -----SSVYSEAGKDKGPPPKRSYREEELRTRGKT-- 1925  
Db 3231 TEQAILBQYVATGEGSOFWATEDATTASTVAIEI-----PQSGWYTVQSDGVTOY 3283  
QY 1926 -----TITAAFNIDVITITROIASDKAREGSQSDSSSSLSSSHRYETPDSAEIVIS 1977  
Db 3284 IAPPGLITVSEIPLTDVYVKEEKQPKK--RSSGAKVRGQYDEMGSMAADDPRNLKKIVD 3341  
QY 1978 PASSAPPOEKLOTY-----OPEVVKANQANDPTROYEGPLHHVYRPOQ-----ESP 2024  
Db 3342 SGVOTDEETADRTYASRRRTKSKSVDTSVOTDDEQDMDPMSRRKARTKGYGDSTA 3401  
QY 2025 SPOQLPSS-----QAEQMGQVPRTHLITLADHICQII-----TQDFA 2064  
Db 3402 EGDTKPKPSKVSSVAVQVAVISVQTEPLGTI-RTPSIRARVDAKVEIKHISAPEKTYK 3460  
QY 2065 RNQVSSOTPOQPTSTQNSPSALVSTPVRTKTNRYSPESQAQSVHHQR-----PG 2116  
Db 3461 GSGLCQCTETDPTD--QSPPYMGATSPPKDK--KRPTPLEIGYSSSHLRADPTVOLAPS 3515  
QY 2117 SRVSPENLVKSRGSRPKSPERSHVSSEPYE---PISP-----POVPVVEKQD 2163  
Db 3516 PKSPKVLVISPISLSPGHLEPAFV---PYEKLPLDDISPOKVLHPDMAKVPASPAKTA 3572  
QY 2164 SLALLSORGAPAEQRDARSPTSISYLPSEFTK-LENTSPMVSKKOEIRKLNSSGGG 2222  
Db 3573 KMQORSMDPKPLSPTADESSRAPFOYSEGFTAKGTSQTSQTKKVKTL-----P 3623  
QY 2223 DSDMAAQPGEIFNLPAVTTSGVSSRSGHGFADPASNGL-LEDIIRKALMGSDDKVED 2281  
Db 3624 NPPPEEASTGTQ---STYSTMGTASRRRMCRTNTMARAKILQDIDREL-----DLVER 3673  
QY 2282 HGVMSPQMGVVPV-----TANTSVTSGETTRREGDPSHSGGVCKPKLISKNSRKS 2336  
Db 3674 ESAKLRKQAELEDEEKEIDAHLRYLEMGINRKEA-----LLKERER 3719  
QY 2337 SPIPG-----QGYLGTTERPSSSVSHSEGDRHQTQWAMEDRPSSTGTOF----- 2383  
Db 3720 AYLOGVAEARDYMSDSEVSSTRPSRVSQHG-----IERPTAPQEFESQFI 3766  
QY 2384 -----PYNPLTMRMLSTPTPTACAPSAVNQAAAPHQONRIWIEREAPPLISAQ 2431



Db 3767 PPQTQTEAQLVPTSPYQYQYSS-PALPTQ-APTPTQSQSHFQQOQLYHQOVSP----- 3819

Qy 2432 YET 2434

Db 3820 YQT 3822

RESULT 20

Q9VID9

ID Q9VID9 PRELIMINARY; PRF: 2897 AA.

AC Q9VID9

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE CG8677 PROTEIN.

GN BEST:LD14959 OR CG8677.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-X., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL: AE003669; AAF53983.1;

DR FlyBase: FBgn0026577; BEST:LD14959.

DR InterPro: IPR001965;

DR Pfam: PF00628; PHD; 1.

DR SMART: SM00249; PHD; 1.

SQ SEQUENCE 2897 AA; 319165 MW; 63900D638B625B38 CRC64;

Query Match 3.78; Score 465.5; DB 5; Length 2897;

Best Local Similarity 18.48; Pred. No. 1.6e-16;

Matches 445; Conservative 302; Mismatches 901; Indels 767; Gaps 97;

Qy 24 SVQYTFPNTRHQOEFAVDPDYRSSHLEVSQASQLLQOQQOQLRRRPSLLSEHFGSDRPP 83

Db 1005 SEETIPRKRNTSETLP-----AIPASNVLCQDPDRHRKRSS-----EDANE 1048

Qy 84 ERRTSYEPFHGPPVDHDSLESKRPL-EOVDSHFORVSAAVLPLVHPLPEGLRASD 142

Db 1049 AFSKESSEPTVPPSAVSEKLRNNEQIQEVED-----PLAMSVKDSLSRSD 1097

Qy 143 AKDPAFGGKHEAPSPISGPGCGDDQNASPSKLSKELIOSMDRVDRDI-AKVEQQILK 201

Db 1098 QSPVPEGSARR-----SGRRGPAVTHSELPPQKTRGARDKMOPEVNAELKQESD 1150

Qy 202 LKKQOQLEEEAAKPEPEKPPVPPVEQKHSRIVQIIYDENRKAEE-----A 250

Db 1151 DEKISTIKSEAKDDPAPES-----ENRKKPEKPIKEEPNEEPK 1191

Qy 251 HKIEGGLPKVELPLYNOPSDTKYVHENIKTNQVMRKLILFFKRRNHARKOREQKIC-- 308

Db 1192 PKVGRGRPKK----REVDTNIIETNDSETPVRQSRRIAQOKIKEAERKQEEVALR 1247

Qy 309 ---QRYDQLMFAWEKVDRIENPRKAKESKTRYYE-----KQPP-----EIRK 351

Db 1248 TMKQELKKKKKA-EKEADPTVLEPSGESESEASEAEAEARNKKKKCPGKDGWSDSEE 1306

Qy 352 QREOQERFORVGORGAGLSATIARSEHEISIIQSLSEONNEKQMRQLSVIPPMMDAE 411

Db 1307 QPESEEEEEPPHYETDPGSPFRSDHEFSP-----ESELEDESQVPM--KR 1352

Qy 412 QRRYKFINMGLMDPMKVYKDRQFMVNVTHDEKEIFEKDFIQHPKNFGLIASYLRKSV 471

Db 1353 ARTYKENADLLEE-----DAEACQCKGSDHPE-----WILLCDT 1390

Qy 472 PDC-----VLYYY-----LTKKNYKALVRRNYKRRGRN 502

Db 1391 PTCNKGYHCSLSPVLYFIPEGDWHCPQOEQLTAALERQEQYDITVAO----- 1441

Qy 503 QOIARPSQEEKV---EKEEDKAEKTEKEEKDEEKDEKSKENTKEDKIDGTAAEE 560

Db 1442 -----KQERILAEQAEERQLEAATLAKDENKFSKEDEDED--RDMVAVGRAEK 1493

Qy 561 TEER-----EQATPRGRKTANSQGRKRGRITRSMITNEAAAAAATAEPPPLPP 613

Db 1494 VKRRGGRGRINRRAAKRG--TRRRRGNSDSSHRKSLGSGSGSGSDSDNSTSFSDS 1551

Qy 614 PPEPI-----STEPVETSRTEEMEVAKGLVEHGRNWAIAKAVGTRK 657

Db 1552 DDEPIYKLRKRRQINVSRYRLNEYDDLINSALKKEMDEVAGNLGRGKDITITE--ADK 1609

Qy 658 SEAQCKNFYENYKRRHNLNLLQOHKOKTSRKPREEDVSOCEVASTVSAOEDE----- 712

Db 1610 EKA-----RRDLP-----EDEVGNKEGKDKQKSKSGSPSSSEDEVPLKR 1654

Qy 713 -----DIFASNEENPEDEVAEPSEDPENATSRGNTPEALEP 755

Db 1655 SNKFKQPPAKKKARKLTLDVSSSEDEHGSDEDFKTSYSDSTQSAS-GDSDSL- 1710

Qy 756 TTETAPSTSLAVPSTKPAEDSEVETQVND--ISAETAQMDYDQOEHAEEGSCVDDP 814

Db 1711 -----AYRPGRGKQKRAARAARERKDRKFVVESEDEDEDQKRRTTKSKKKKDD 1765

Qy 815 PAYKADSVDEVVRVPENHASKVEGDNKTRD-----DRASEKVE-----PRDELVAAQOI 866

Db 1766 DYETETEDDD-----DNELS-----DNVDSADLDCDDTTSEEDGANWSSKKKTVAAKS 1817

Qy 867 NAO-----RPEPQ-----SDNSSATCSADEVDGEPERQRMPPMDSKP 905

Db 1818 NSSGGIARKSPKLLKATQAEKKVKRLEYSDDDISES-DLEEDDDDDDEDEEGVPLSGK 1876

Qy 906 SLLNPTGSILVSSPLKPNPLDLQLOHRAAVIPWVSCVTCNIPIGTPVSGYALYQRIK 965

Db 1877 SGKQPR-----SQPLKPT-----ASSTLTGKGK-----RGKAKKKKQVS 1911



QY 966 AMHESALLEEQROR-----QBQIDLE-----CRSTSCGTSKSPNREWEVLQ 1008  
DB 1912 SEEDGAASDDTRTRRRYAYIEDDDSDGKIPGVHRPDTPEERQKFQIQEEI-- 1969  
QY 1009 PAPHOLITNLGVELPTTRPPPLIPSSKTTVAASEKSFIMGSSISQGTPTGYLTS 1068  
DB 1970 --KRLMAEKNAEAKIAAT-----PRLTPLKSGVTASEK-----RTPG----- 2005  
QY 1069 HNAQSYTOTETPKPSVGSISLGLPROQESAKSATLPIYIKOEESPRSONSQPBGILLVRAQH 1128  
DB 2006 -----KASGDSLSLTVLSVIRO---AKVLIDIDL-----QR 2034  
QY 1129 EGVRGTAGATQEGSITRTPTTSKTSVSPSLRGSITQGTALPQTGIPTEALYKGSIS 1188  
DB 2035 KGETIGDLDVDE-----SELDDAELPD-----DLPED-----MEDAIA 2068  
QY 1189 RMPIEDSDPEKREAAKSHGVYIEGKSHILSYDNINKNAREGTSRPTAHEISLKRSEY 1248  
DB 2059 RWVEEE---EQFSAEVAAR-----ELPGAEE----- 2091  
QY 1249 SVEGNIKQMSMRESPVAPLEGILCRALPGSPHSDLKERTVLSGSIOMGTPRATTESF 1308  
DB 2092 -----VLRTPSKSQ-----TSRVMPSPANSPS 2118  
QY 1309 EDGLKYPKQ-----IKRESPIRAFEAGITKGPYDGITTIKEMGRSIEHPRQDILTQES 1364  
DB 2119 TSQLQEPHRKRLPMTMPHLLRHQFPISAGSPHSASIV----- 2158  
QY 1365 RKTPEVQOSTRPIEGSISQGTPTKFDNNSGOSAIKHNVKSLITGPSKLSRGMPPLEIVP 1424  
DB 2159 --PPHAAGMHPMLQRHLSQTV-----PPOAMHLLQNALSP-----LQGP----- 2198  
QY 1425 ENIKVVEGKEDVKAGETVRS-RHTSVSSGSPSVLRSTLHEAPKAQSLPGIYDDTSARR 1483  
DB 2199 -----LGCNRY-----GSGPNSAQHLPVMSPSAAAAAHLMOSAVAS-----ATARP 2242  
QY 1484 TPVSYQNTMSRGMNRTSDVTIPPKNSTNHERKSTLTPTQRESIPAKSPVPGVDVYVS 1543  
DB 2243 VETASGNPASDPK-GRRRKKVTLRDQLQKQTAATAVTAATSTTPGSAP--SEKYKA 2298  
QY 1544 HSPDPHRRGSTAGEVYVSHLPTQLDPAMPFRALDPAAYLAFORQLSPITPG-YPSOYQ 1602  
DB 2299 QLFKPHEDAAPSAPASQAVITRMSPLP-----PAHG-----RNHGPPSGLYPSSAD 2347  
QY 1603 L---YAMENTROT-----LNDYITSOQMOVNLRPDVARGLSPRE---OPLGLPYPATR 1650  
DB 2348 LARFYGVANQOPIPAVPGSRSPSTSGPPHLLRPQPPGLPPHSLRLPYGPPPLR 2407  
QY 1651 GIIDLNNPPTIL-----VPHPGGTSTPMDRITYIPGTQITFTPPRPNYSASMS- 1700  
DB 2408 GSGPPTSTPTTNSRPAYLHGAEHGGGSPGPMGV-FSSGP---PPARHATPHLNPY 2462  
QY 1701 -----GHPTLHAASAERERERERERERERERERERERERERERERERERERER 1750  
DB 2463 RAPPYIGNPNT-----SPRLGAPGTCGMRPGAVDVVAGPRGYSP 2502  
QY 1751 HGYVRSPSVYRTOETMLQRPVSFQGTNGTSVITPLDPTAQLRIMPL-----PAGCP 1803  
DB 2503 YGYTPPPPP-----LSTPSHAHAATSVSVISAPHTLTNTNHSVPTLTHGKTPPOQTP 2553  
QY 1804 SISQGLPASRYNTAADALVALDAASAPQMDVSKTKESKHEAARLEENLRSRAAASVSEQ 1863  
DB 2554 TQSSGPP-----PAAAPPTITSETSHKPPPLASV---ITSKLTITLEA 2594  
QY 1864 QOLBQKLTLEKRSVQCLYTSATSPSGKQPHSHSVVYSEAGKGGPPPKSYBEELRTRG 1923  
DB 2595 YPIRKSPI-----AVADVSGPAEPTRSAP-----IAEEDSG-----SAHD-----TRA 2634  
QY 1974 KTTITANFIDVIITQASDKDARERGSSQSSDSSSSLSHRYETPSDAIEVSPASSPA 1983  
DB 2635 PSSATGTAIV-----GEPFSGSGNAQWAAH-----CTGSPT 2663

QY 1984 P---POEKLQTYQEVVKANQAEENDPTRQYEGPLHHYRQBPQSPQOQLPSSQAEGMG 2040  
DB 2664 PRELGQSKLELEQEQ--SKLEREQEPSKL-----ELELEQEQSKLERAQEREGPG 2711  
QY 2041 QVPRTHRLITLADHLCIQLITQDFARNQVSSQTPQO---PPTSTFFONS----- 2084  
DB 2712 R-----EVTRELAQEQSKLELGQEQSIPETAQEQNTLGREQCCLQEAD 2756  
QY 2085 --PSALVSTPVRTKTSNRYSPESQAQSVHHQR-----PGSRVSPENLVDSRGRSPCK 2135  
DB 2757 TPKRVLVREQSELEGLGERSKQEQVQERSKRELEPECPGREASTQELAQOSKLELGQ 2816  
QY 2136 SPERSHVSE-----PYEPISPPQVPPVHEKQDLSLLLSQRAEPAEQRNDARS 2184  
DB 2817 EQSKLELEQEQNKQELGCLQEAGTPQVLVPSRSLGLGERSKQEQSQEQSTQELE 2876  
QY 2185 PCSISYLSFFTKLE 2199  
DB 2877 PAQISPPPERVADKLE 2891  
RESULT 21  
Q9VH10  
ID Q9VH10 PRELIMINARY; PRT; 3111 AA.  
AC Q9VH10;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG3996 PROTEIN.  
GN CG3996.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20195006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abriel J.F., Agbayani A., An H.-J., Andrews-pfankoch C., Baldwin D.,  
Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berham B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Garg N.A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclev J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,



















Db 2400 IVTHVDAL---TWVEAAARRS-----NGLPATKCAIAPP-----VPPKPSQIPTGL 2446  
Qy 1019 -----PEGVRLPTTRTPPELIPSSKTIVASEKSPFIMGSGISQGTGGVTLTSHNQAS 1073  
Db 2447 VTHRPEAIKPIA-----PXPAPV-QIPVTTQKP----- 2475  
Qy 1074 YQETPKPS-----VGSISGLPQROE-SAKSATLPYIKOEFSRPSQSOPE----- 1120  
Db 2476 -TDCPKPTGLSLTSLTMSLNLVTSADYNVPSPTSP-----LSPHSNKSPPRYKSLMDT 2528  
Qy 1121 -GLLVRAQHEGVVWGRTAGAIQEGSITRTPTSK-LSVESIPSLRGSITQGTPALPQGTGP 1178  
Db 2529 YVITLTPSEPQPTDSSAAQAITSPLGSPKDLVSLTETFSV-----VPPMTSTETP 2581  
Qy 1179 TEA-----LVKGSISR-MPIEDSSPEKREGAASGHVYI 1212  
Db 2582 SASQPTLYTSGALGTFVTPAVTASLFTVPTLSLQFLPAEASRPEVSASVAP----- 2636  
Qy 1213 EGKSGHILSYDNINKNAREGTRSPRTAHEISLKRYSVESYEGNIKQGMMSRESVPVSAPLEG- 1271  
Db 2637 -----SVAPRSVSIPIPEPLALDRHQYKENGKL-----PLIGD 2670  
Qy 1272 -LICRALPRGSPHDLK--ERTV-LSGSIMOGTPRATTESTEDGLKYPKQIKRESPPPIRA 1327  
Db 2671 AIDLRTIPK-----SEVKYTERKMDLSASAMDVKQTTANEV-----YRQISAVQPSIIN 2721  
Qy 1328 FEGATTKPKPYDGIITIKEMGRSIEIHPQDILTOESRKTPEVWQSTRPIEGSISQCTP 1387  
Db 2722 LSASSLGT-----VTMDS-KTAVAVTCT-----DTT 2748  
Qy 1388 IKFDNNSQSAIKHNKVLITGPKSLRGMPPLEIVPENIKVVERGYEDVKGAVTBSR 1447  
Db 2749 IYTTCTESQVGEHAVTS-----PLQL-----TTSK 2774  
Qy 1448 HTSVVSSGSPVLRLTLHEAPKAQLSPGYDDTSARTPTVSTQNTMSRSGPMNRTSDVTI 1507  
Db 2775 HTELP-----YRKPSQAFPTIRDE-----APINLSLSPSAQAQAVTLAVTKPTV 2818  
Qy 1508 PNKSTNHERKSTLTP-----TQRESIPAKSPVPGVDVPSH-----SPEDPH 1550  
Db 2819 PPVGVNWTSTLSQGVADGEVDLSTSKSHRTVVTHDESTSNVTKLIEDDEKPDV-- 2876  
Qy 1551 HRGSTAGE-----VYNSHLPTQLDPAHPFHR-----ALDPAAAA-----YLFQR 1589  
Db 2877 ---LTAGRAVCCDMVY-----TLPEGRSCTAQAPATTLPEDRFGYRDDHYYQDR 2923  
Qy 1590 QLSPTPGY-----PSQ-----YQLYAMENTRQTLINDYITSQOMOVNL---RPD 1630  
Db 2924 --SGPYGRGIGGKPKMSDNLNPEAGHFFYKSKNA-----FDYSGGTGAADVLTSGRVS 2976  
Qy 1631 VARGLSPREQPLGLPYPATRGITDILNPPITLVPHPGCTSTPPMDRITYIPGTQITPPP 1690  
Db 2977 TGEVMDYSKTTG-PYPETROVISGV-----GISTPQYST-----ARLTPPP 3017  
Qy 1691 RP-YNSASMSGPHTHLAAAASAEERERERERERERERERERERERERERERERERERER 1746  
Db 3018 GPQYGVGSV-----LSSNGVYSSVATPIPSRFA 3047  
Qy 1747 ---RGS--HGVRS-----PSPSVRTQETMLQO---RPSVFOQTNGTS-----VITPLDPT 1790  
Db 3048 ITTOPGSIFSTVTRDLSGIPITDAMTSLSLAQHOSOPMRSPFITGASETIADVIGDIN 3107  
Qy 1791 AQLRIPLPAGGSPISQGLPASYRN-----TAADALAALV-----DAAAASAPQM 1834  
Db 3108 ASLQITIM-----ETLPAETMDSVPTLTITASEVSEVGEESTLLIVFDEDKQQOOL 3159  
Qy 1835 DVSK-----TKESK 1843  
Db 3160 DLERELLELEKIKOORFAEELEWERQEQIRFEQEKIMVQKKLELOSMQKHLLYQOBBE 3219  
Qy 1844 HEAAAL--EENLRSRAVSEQQOLEQKT-LEVEKRSVQCLYTSAPPSGKPPQHS----- 1896

Db 3220 ROAFMMRQETLAQQQLQLEIQIQQLOOQHOOLEBQKLRQIYQYNDPSGTSSTPQTTEQ 3279  
Qy 1897 SWVYSEAGKDKGPP-----PKSRYEEELRTRGKT-----TI 1927  
Db 3280 AILEGOYAATSGSQFWATEDATTASTVVAIEIPQSQGWYTVQSDGVTOYTAPPGLITV 3339  
Qy 1928 TAANFIDVITRQ-----IASD-----KDARERGSQSDSSS-- 1959  
Db 3340 SEIPLTDVVVEEKQPKKRSSCAKRGVQYDEMGESVADDPRLNKKIVDSGVQTDDEETAD 3399  
Qy 1960 -LSHRVETSDALEVSPASS-----PAPPOEKLOTQYQEVVVKANQANENDPTROYE 2011  
Db 3400 RSYASRRRTTKKSDVTSQTDDEQDEWDMPSRSRKRARTGK---YGDSTAEGRDK-- 3453  
Qy 2012 GPLHHYRQOQSPSPQOOLPPSSOAGMGQVPRHRLITLADHICQI1-----TQDFAR 2065  
Db 3454 -PL-----SKVSSVAVQTVAEISVQTEPVGTI-RTPSIARVDKAVEILIKHISAPEKTYG 3507  
Qy 2066 NOVSSQTPQPPPTSTFQNSGALYSTPVTRKTSNRYSPEQAQSVHHQR-----PGS 2117  
Db 3508 GSLGQCTETDSDT---QSPPYLGATSPPKDK--KRPTPLEIGYSSSHLRADPTVQLAPSP 3562  
Qy 2118 RVSPENLVDKSGSRPKSPERSHVSSEPYE-----PISP-----POYPVVEHKQDS 2164  
Db 3563 PKSPKVLVSPISPLSPGNALPEAFV---PYEKLPLDDISPOKVLHPDMKVPASPKTAK 3619  
Qy 2165 LLLLSORGAEPAEQNDARSFGSISYLPSPFTKLENTSPMVKSKQEIFRKLNSGGGDS 2224  
Db 3620 MMQSRMSDPKPLSPADESSRAPFOYSEGFTYKGSQT--MTASGTQKKVKRTPLPNPPPEE 3677  
Qy 2225 DMAAAQPGTEIFNLPAVTVSGSVSRGHSFADPASNLG-LEDIIRKALMGSDDKVEDHG 2283  
Db 3678 ---VSTGTQ---STYSTMGTFASRRMCRNTMARAKILQIDREL-----DLVERES 3723  
Qy 2284 VMSQPMGVPG-----TANTSVTVSGTTRREEDGPPSPHSGVCKPKLISNSRKSXP 2338  
Db 3724 AKLRKKQAELEDEEKEIDAKLRYLEMGINRKEA-----LLKREKREKRAY 3769  
Qy 2339 IPG-----QGYLGTFRPSSVSVHSEGDYHROTQGAWEEDRPSSTGTQF----- 2383  
Db 3770 LQVADRDYMSDESEVSTRPSRVESQHG-----VERPTAPOTESQFIPP 3816  
Qy 2384 -----DYNPLTMRMLSTPTPTIACAPSANVQAAPHQOQNRWIREPAPILSAOYE 2433  
Db 3817 QOTQEAQLVPPPTSPYTOYQYSS-PALPTQ-APTPYTQOSHFQOQTLYHQVSP-----YQ 3869  
Qy 2434 T 2434  
Db 3870 T 3870  
RESULT 24  
Q9VC00 PRELIMINARY; PRT; 2768 AA.  
ID Q9VC00  
AC Q9VC00;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CGI3648 PROTEIN.  
GN CGI3648.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,







```
QY 1344 IKEMGRSHEIPRODILTO-----ESRKTPYVQSPRPIEGSISQGTPIKFDNN-----SG 1395
Db 1465 TE-----ADISTEPSAWEKEASCETSESDNEIDAGASSTPPVPSADEDKTST 1513
QY 1396 QSAIKHNKSLITGPKSLRGMPPLEIVPENIKVVERGYEDVKAGETVRSRHTSVSSG 1455
Db 1514 EKTVEADDDFTTAVPLAGDEESNLPKLPQDI-----FEB-----EAPVAVTTAA 1558
QY 1456 PS-----VLRSTLHEAPKALSPGIIYDTSARTPVSYNMT-----1492
Db 1559 PSKDGEQKPVVEEKPIEDGQKPIEDETS---TPTSSENEIPESDRATTIAPKEEPS 1615
QY 1493 --SRGSPMMNRSDVTIPNKNSTNHERKSTLPTQR-----ESIPAKSPVPGVDV 1541
Db 1616 EFTSGAP---TKDEPAEPSTDAPESDSKETPESEVPTTVAPAGEKIPTSITIDEET 1671
QY 1542 VSHSPF---DPHHRGSTAGEVYWSHLPTQLDPAMPFHRALDPAALYLFQRLSPTPGYP 1598
Db 1672 ATSAPVAKPDEDEKETSTEI-----PTDA-----PASSE---EDENSTDDQIP 1712
QY 1599 SOYQLYAMENTQITLINDYI-----TSQOMVNLRPDVARGLSPREQPLGLYPATR 1650
Db 1713 SEVPEKKPETAQTPPEGDIVGATAAPTSDV-----PPVQR---LPEEVLAIEIPQSTE 1765
QY 1651 GIIDLTMNPPTILVPHPGTSTPPMDR---ITYIPGTOITPPP-----RPNYSASMS 1700
Db 1766 -----TGIKQOEDTAAPSIDREKPYTEIDEATTVPALSEKDEKTEEEKPVE 1815
QY 1701 GHPTLAAASAERERERERERERE-RIAAASDLYLRPGSEQPGRGSHGYV-----1754
Db 1816 KQPT--GEEPSBEEKEKPIEQDVSTEGVPSTEASEAGSTESSEV-KPSTGEVAEKPE 1872
QY 1755 -RSPS-----PSVRT-----QETML---QORPSVF-QGTNG 1780
Db 1873 DKQPSSTAQAVETIPELSTELPAQDGKPTSEAPVDSDETSAPSKIPSVSGEEVEG 1932
QY 1781 TSVITPLDPTAQLRIMPLFA--GGPSISQGLPASRY-----NFAAD-----ALAAALVDA--1828
Db 1933 PVTVTASQAEEDELKTPAEPSEPSSTDKVPETEYQKPEDETKADETPESVTVQSDVATS 1992
QY 1829 ASAPOM--DVSKTKESKHEAARLENLSRANAESVQOQLQK-----TLEVKRSVQCQLY 1882
Db 1993 TSAPVAGGDIKDEQATTSAPPEEIEKPTIAPAAEIPQSEKEPVDEQEVESGKTATPA 2052
QY 1883 TSSAPPKGPPOPHSWYSEAGKDKGPPKSRYEELRTRGKTTITAA-----NFID 1934
Db 2053 ESDGQPIDEIAPATSGPIDEA--STAAATK-----EE-----STTVASAASPAVHDEIKD 2101
QY 1935 VIITROIASDK-----DAREGSSDSSSSLSLSSHRYETPS 1970
Db 2102 VTTTPVADAEKEVAAPQDETSTSIDVSTDSPTAQDDEKQDKTEAPVATTVSSPTADSA 2161
QY 1971 DA----IEVISPASPAPQEKL--QTOPEVVKANQANDEPTROYEGPLHIHYPQOE--2022
Db 2162 DSSTPEVPSPEVDITKPMDDIMSQTAPHT--ADGAASTSTEDQOAPVTVSPQDAEK 2219
QY 2023 ---SPSPQ--QOLPPSSQAEQGVPRP-----ADGAASTSTEDQOAPVTVSPQDAEK 2219
Db 2220 TPVSPAPQDSKTPSSAPQADEIPATATPLDDNKIPATVAPQTDGCVATAAPLDEDK 2279
QY 2048 LITLA-----DHICQIITQDFARNQVSSQTP-----QOPP---2077
Db 2280 IQTTAAPLDEEKIPSTAAPLDEKIPAPVSPVFDVPSSEKPAVSEYDGEESTEPVHD 2339
QY 2078 --TSTFQSPSALVSTPVYTKTSNRYSPESQAQSVHGORPSRSPENLVOKSGRSRPGK 2135
Db 2340 VETSTDEPSTDAKLPTSPATPSESATPEAEIV-----PETAAPLEKEVPEK 2389
QY 2136 SPERSHVSESE-PYEPISPPQVPVVEHEKQDLSLLLSORGAEPAEQNDARSFGSISYLPSP 2194
Db 2390 ATEQPELEKETPERKATEQPEL-----EKE-----TPEKATEQP--2422
```

```
QY 2195 FTKLENTSPMKVKOEIPLFRKLNSSGGSDMAAOPGCTEIFENLPAVTTSGSVSRGHSF 2254
Db 2423 --ELEKETPEKATEQPELEKEV-----TKRATQPE-----SVDEKTTT-----2459
QY 2255 ADPASNLGLEDIIRKALMGSDFKVEDHGVMSQPMGVVPGTANTSVVTSGETRREEGDP 2314
Db 2460 -----EPVVKPSL-----DSTEDEESVESEESADKKDKNKE--TEEDTKKHEEP 2504
QY 2315 SPHSGCVCKPKLISKNSRKSKSPIGQGY-----LGTERPSSVSSVHSGDYHROT 2367
Db 2505 E-----VPVVSEI-PQPSSEAVPTTGHPLFPHLASSTTTTTPPAVDVDRVGE-----2548
QY 2368 GWAWEDRPSSTGCTQFPYNPLTMRMLSTPTPIACAPSAYVQAAPHQONRIRWEREAPL 2427
Db 2549 ---EDEENTTVK-----LSSSTTTTSPVTSAPSTTTTVAOQQO-----PITPPPY 2593
QY 2428 LSA-QYETLSDSD 2440
Db 2594 GHAPYEDEYDEE 2607
AC Q9UHA8 PRELIMINARY; PRT; 2296 AA.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE SPLICING COACTIVATOR SUBUNIT SRM300.
GN SRM300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA Rosonina E., Sharp P.A.;
RT "The SRM160/300 splicing coactivator subunits.";
RL RNA 0:0-0(2000).
DR EMBL; AF201422; AAF21439.1; -.
DR InterPro; IPR002965; -.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 2296 AA; 251962 MW; 17C0BD4EAL0A9CF9 CRC64;
```

Query Match 3.6%; Score 453; DB 4; Length 2296;  
Best Local Similarity 19.6%; Pred. No. 5.7e-16;  
Matches 504; Conservative 288; Mismatches 979; Indels 796; Gaps 114;

```
QY 44 RSSHLEVSQASQLLQO-QQQQLRRRPSLLSEFFHPGSDRPOERRTSYEPFFHPGSPVDHD 102
Db 305 REMLSVNOVLPSYQRASSPETATKQPS-----SPYEDKDKKKEK-SATRPSFSP-----354
QY 103 SLESKRPRLEQVSDSHFORVSAAVLPLVHPLPEGLRASADAKDPAGFGKHEAPSPISG 162
Db 355 -----ERSSTGPEP---PAPTLLAERHG-----GSPQPLATTPLSQ 388
QY 163 OPCGDDQNASPSK-----LSKEELIQSDMRVDREIAKVEQOILKLLKKQOQLEEAAPK 217
Db 389 EPVNPPEASPTDRSPSPKPEKLQSS---SSESPSPQPTKYSR-----HASSP 438
QY 218 EPEKVPSPPPVQKHRSITVQIYDENRK---KAEAAKIFEGGLPKVELPLYNOPSPTK 273
Db 439 ESPAPA---PAPGSHREISSPTSKNRSHGRAKDKSH---SHTPSRMGRSRSPATAK 491
QY 274 VYHENIKTNQVMRKLILFFKRRNHARKQREKICQRYDQLMEAWKKVDRIENNRKA 333
Db 492 RGRSRSTP-----TKRGHSRSRSPQ---WRRRSQAQRWGR-----SRSPQRRG 532
QY 334 KESTREYIEKQFPEIRKQREOQERFORVGORGAGLSTATARSEHEISEIIDGLSEQENN 393
Db 533 RSRS-----PQRPGRSRRNTQRRGRSARRGRS-HSRSPATRGRSRSTP 578
```















RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weissman G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
CC DOMAIN.  
DR EMBL; AE003473; AAG22226.1; -;  
DR InterPro; IPR001452; -;  
DR InterPro; IPR001777; -;  
DR InterPro; IPR003006; -;  
DR Pfam; PF00018; SH3; 1.  
DR Pfam; PF00041; fn3; 5.  
DR Pfam; PF00047; Ig; 11.  
DR PRINTS; PR00014; FNTYPEIII.  
DR PROSITE; PS50002; SH3; 1.  
KW Repeat.  
SQ SEQUENCE 6815 AA; 779559 MW; A4E244001A4EBA01 CRC64;

Query Match 3.6%; Score 450.5; DB 5; Length 6815;  
Best Local Similarity 17.7%; Pred. No. 3.4e-15;  
Matches 430; Conservative 376; Mismatches 906; Indels 713; Gaps 104;  
Qy 14 STEQKRYPHSVYFPNTRH-----QOEFAVPDYRS-SHLEVSQ--ASQLLQOQOQOQ 64  
Db 2934 SSEQPK-QPSTEQYELSVTEHDLKPEEKPTFTVQIQSTNVEETKDDTGKVKHQVTKR 2992  
Qy 65 LRRRPSLLSEFH-----PGSDRQERT--SYEPHFGSPVDHDSLESRP--RLEQVSD 116  
Db 2993 MLRRPAGEGEIEIEVVRDDQPEAEITVEYE-----PEPVNQDE-KPKPEKKTKRKK 3046  
Qy 117 SHFQVSAVLPLVHPLPEGLRASADAKDPAGFGKHEAPSPISGQCGDDQNASPSKL 176  
Db 3047 DDHDIYQKLELETPKTE-LEKYEKIEFEPIV--KDKPLDSPI-----DVLDESPEV 3097  
Qy 177 SKEELIQSDMRVDRIAKVEQOILKLLKQOOLEEAAKPEPEKPVSPPPVQKHRSIV 236  
Db 3098 QKDKKSRSTKVPNEETPVQYAKV---NVVEEA--PEQELPVQILEVKKPVEVDVK 3151  
Qy 237 QIYDENKKAEEAHK-IFEGLGPKVELPLNQPSDTKYHYHENTKTNQVMKKLLILFKR 295  
Db 3152 EVITEDGKPVQEKTKRLVKLIGPEQTFE---KITMIESDNDSDVTVI-----3197  
Qy 296 RNHARKQREQICQRYQDOLMEAWKKVDRIENNR--RKAKESKTREY----YEKQPE 348  
Db 3198 -----VDEEPEIASP-QSIEEHPQOSKEKAPKPKTKVKKKDDLSYVKKLIEEELPK 3251  
Qy 349 IRKQREOQERFORQORGAGLSATIAARSEHETSEIDGLSEQENNEKOMRQSLVIP-----404  
Db 3252 V-----DLEKYEKEMPEKPKVLTVS-----DSIPEEPKPKDS-QPISVLPTDK 3295  
Qy 405 -----PMFQAEQRVRVFINNGLMEDPMKYKDRQFMNVWT-----441  
Db 3296 PKTKTKPTKPTEDTDOQ-----VPDEETTTVDITDIPELTPTQTAQEDTATAQ 3346  
Qy 442 -----DHEKEIFKD-----KFIOHPKNFGLIASVLERKSVPD-----CVLYYLTKK 483  
Db 3347 ITPSAEKEKSDQDDTKDTIQTKTVKHKTKPDTQKSVSETSELPEVHKDYQISIIHEELVEE 3406

Qy 484 NENYKALVRRNYGK--RRGRNQIA-----RPSQOEKVEE--KEEDKAEKTEKKKEE 531  
Db 3407 EQPEKILEVRVIDEVAEVEESQPIVEEVEDEEPOPATETEVEDTKPSKKKKVVKKTD 3466  
Qy 532 KKDEERDEKEDSKENTK-----EKDKIDGTAE-----ETEERQ-----A 567  
Db 3467 DHDELIKMLQEOIEKTELEKYEIEFDVPPKKLPEFAALPEPIKIERKEQKPTKVITLDA 3526  
Qy 568 TPRGRKTANSQGRKGRTRTSMTEAAAA-ASAAAAAATEEPPPLPPPEPIST--EPE 624  
Db 3527 TDVPKTVKLKSKKEKPAEELTVQLPFRKARMLVVEYPPAPLIPKTTDIGAINDGE 3586  
Qy 625 TSWTEEMEVAKKGLVBEHGRNWAIAAKMVGTSQAQCKNFYFNKRRHNDLNLQHQK 684  
Db 3587 LSRNIEEAELK-----FPHKPT 3605  
Qy 685 KTSRKPREDVDVSCQESVASTVSAQEDD-----TEASNEENPEDSEVAVK-----732  
Db 3606 KKIKKIKDDLEKVELEKYEKYSSEEEPEEPTPKPKPKAPKEEKQEDVKLKGKKK 3665  
Qy 733 -PSDSEPNATSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDSVETQV-----NDS 787  
Db 3666 KPEEAPENVTLKNIPQKQVEEVEEL-----KOKPKEVEIVEQTKPKDGE 3714  
Qy 788 ISAETAQMDVDOQHSAAEGSVCDPP-----PATKADS-----821  
Db 3715 FWEPFEPSEFDRPEYVPDELEQIEHPEIPEKVKKPSKTKYKPKDKSKSEPETIVSEIVA 3774  
Qy 822 -----VDVEVRVPHNASKVEGDNTKERDLDRASEKVEPRDEDLVVAQINQAR 870  
Db 3775 GVPKEEAIPQDVKFRKPERDAPETSEIKLRVPQAQSKDENPDEQALVTPK---AEE 3831  
Qy 871 PEPSDNDSSATCSADEVDGEPERQRMFMDSKPSLNPPTGSIILVSSPLKPLDLPLQL 930  
Db 3832 PIPOEIDKA-----IDDEKK-----PKSRPKVPQKEQETAKE--EPPEEVSVK 3876  
Qy 931 QHRAAVIPPMVSCPCNT-----PIGTVSGYALYQRHIKAMHESALLEEQRQORQ---E 981  
Db 3877 EEEALVDKPIETKPKDKVKKEKPKAPVS-----EVVIEEPEKPEEVE 3923  
Qy 982 QIDLECRSTSPGCTSKSPNR-----EWEVLQAPHQLITNLPEGVRLPTTRTPRPPPP 1035  
Db 3924 EIPVEYKITTTLVLEPADAPKQHVQVIDFERQETEEVI---EEKV---VTRKKKPKQ 3977  
Qy 1036 LIPSKTIVASEKPSFIMGSGISQGTGTYLTSHNQASYTOETPKPSVSGISLGLPROQE 1095  
Db 3978 QPEFEVTLKEPK-----EQIQPDVVSABSLPIEE-4009  
Qy 1096 SAKSATLYIKQEEFS-----PRSQNSQEGLLVRAQHEGVVGRGTAGAIQEGSITRGTPT 1150  
Db 4010 -----PEQKPEQYEVKELKITQITPEEPNDVQI-----AVKEKVKTK--PV 4047  
Qy 1151 SKISVESIPSLRGSITOGTTPALPOTGIPTEALVKGSIISRMPIEDS--SPEKREBAASKG 1208  
Db 4048 KVKREDKI-----VVVEAEEKQVPEETIVEVEKQEKKKSEK 4085  
Qy 1209 HVIYEGKSGHLSYDN--IKNAREGTRSPRTAHEISLSKRSYESVEGNKQ-----MSMR 1261  
Db 4086 PKSEFKISETQSEERKPEIAEAEAPETPKVKKVAEKDFSYETLUKEDEKVIITVD 4145  
Qy 1262 ESP-VSAPLEGLICRALPRGSPHSDLKERTVLSIMQGTPRATTE-SFEDGLKYPKQIK 1319  
Db 4146 DQPEEAPVE-VVFKKKP-----KEPEAVEAFVMTPEKIVEETSVETAIKQKTKK 4196  
Qy 1320 RESPIRAFEGAITKPKPYDGIITIKEMGRSIIHEIPRODILTQESRKTPFVVQSTRPIIE 1379  
Db 4197 -----PKKD--EEEAQLAIKVVESAPVAE 4219  
Qy 1380 GSIQGTPIKFDNNSGQSAIKHNKVLITGPKSLSRGMPPLIIVPENIKVYVGRGYEDVK 1439  
Db 4220 EVFSEA-----PESKIVVEEIVAEKPEFTIR 4247  
Qy 1440 AGE-----TVRSRHTSV-----SSGPSVLRSTLH---EAPKAQ 1470



```
Db 4248 VSESEPKPEPSVEQTVKKRKPSTVFADEPATEIVIKESKPAEVVTDHAHIKTKPKKK 4307
Qy 1471 LSPGIYDTSARRTPVSYQNTMSRSGSPMMRTS-DVYTPPNKST-----NHERKSTLPT 1524
Db 4308 VT-----DVEAEELIKITEEYVQPIILEEVESEVETETKKTAPVVEEKYKIGIKET 4362
Qy 1525 QRESIPAKSPVGVDPVYVSHSPDFPHRGSTAGEVYVSHLPTQLDPAMPFHRLDPAAAA 1584
Db 4363 EPEK-PAEAVIEEPEPVTE-----PIEA-----4386
Qy 1585 YLFORQLSPTGPGYQOYLAMENTROTILNDYITSQMQVNL-----RPDVARGLSPREPQ 1641
Db 4387 -----PKPEVEEHKVRVTEETPRELVEEIVKVRIRKKPKPEI-----KEEP 4432
Qy 1642 -----LGLPYP-----ATRGIIIDLTMPP-----TILVPHGGYSTPPMDRITYIP 1682
Db 4433 EAEVTVSTPKPVEEVEATSIIVPEQTEEEAALDKITII-----EBETPPELVQEI 4487
Qy 1683 GTQITFPRPYNSASMGPHPTHLAAASAERERERERERERERERERERERERERERER 1742
Db 4488 EIEIVEEKAPEE-----QPTDFTFKDKSEKKTVEELPEQVTIQKKKK---APVP 4538
Qy 1743 EOPGRGSHGVYRSPSVR-TQETMLQORPSVFOGTNGTSVITPLDPTAQLRIMPLPAG 1801
Db 4539 EVVEEPEAEFLVKPTPVQVTEEAKI-----TKSKPKVKEEAAALKV-----4583
Qy 1802 GPSISQGLPASRYNTAALAAALVDAASAPQMDYSKTESKHEAARLEENLRSAAVS 1861
Db 4584 --TITEEIPTE--PEVQELIEIEIEIEIEEKPAEYVIEKESQPEA-----4624
Qy 1862 EQOQLEQKTEVEKRSVQC--LYTSAPSPGPKPQPHSSVYSEAGKDKGPPPKSYEBEL 1919
Db 4625 ---VEDKEVSLPKKKKRAPIVEEPEAEITLKPVKSEEVQEEAKIVKKPKKI---DEV 4677
Qy 1920 RTRGKTTTAAFI--DVIIITQIASDKARERQSSDS-----SSLSHRYETPSDAIE 1974
Db 4678 AVADELTVKVEEVEVPEIVEEVEIEFEIKKKPEPEPIDVDAIVKLKPEPVDAD 4737
Qy 1975 VISPASSAPPOEKLOTYQ-----PEVVKANQANDPTROYEGPLHRYRPOQESPS 2025
Db 4738 VV--AEVILPKPAKTEVTEEEESVDVLPKEKKVTEESDQTVQL-----KKKK 4785
Qy 2026 POQPLPPSQAEGMGVPRTHRLITLADHICQIITQDFARNQVSSQTPOQPTSTFQNSP 2085
Db 4786 PKPVEEADELKQTVVEERPVIEEE--EIIIEAV---VIRKKPKKPEPTVEDLE 4839
Qy 2086 SALVSTPVRTKSNRYSPESQASQVHQRPGSRVSPENLVKSRGSRPKGSPERSHVSSE 2145
Db 4840 ETEFSLSPKKPHTINEGVE-EAATVLLKRP---VRPTTL-DEAAAELSIKRQEEYE 4894
Qy 2146 PYEPI-----SPPOVPVVHKKDS 2164
Db 4895 DIEEFVSVQQRKPKPLQITEEDEEA 4919

RESULT 28
Q9N541
ID Q9N541 PRELIMINARY; PRT: 4900 AA.
AC Q9N541:
DC 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE Y2C2A.1 PROTEIN.
GN Y2C2A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
```

```
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RL Science 282:2012-2018(1998).
RN [2]
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006731; AAF60483.1; --
SQ SEQUENCE 4900 AA; 551244 MW; D837F46317C902D5 CRC64;

Query Match 3.6%; Score 449.5; DB 5; Length 4900;
Best Local Similarity 19.6%; Pred No. 2.5e-15;
Matches 584; Conservative 388; Mismatches 1098; Indels 917; Gaps 139;

Qy 24 SVQYTFPNTRHOOEFAVPDYRSSHLE-----VSQASQLLOQOQOQLRRRPS 70
Db 181 STENEITNTSQTANIVDPGSDIDSEENKLAALKVIEESILKASQEAASDNHEKERSAT 240
Qy 71 LLSEFHPGSDRPOERTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQV-----122
Db 241 SKADY-----ERSFDQDVTYERSPLLEPSEEPVMSKEPELTQEEIDHIAQSLAEQSS 296
Qy 123 --SAAVLPLVHLPEGL-----RASADAKDPAFGGKHEAP---SSPIS 161
Db 297 FQOASTIPDRPPLPARLPTVDEPIVISEQHEEDRSATSGADYERSFDQEVYERSPLL 356
Qy 162 GQPCDDQNASPKLSKE-ELIQSMRDVREIAKVQOILKLLKKQOOLEEAAKPEPE 220
Db 357 -EPCEE-----PVMSKEPELTQ--EIID-HIAIQS-----LAEKSSPEQASTIPDRPP 402
Qy 221 KPVSPPPVEQKHSRIVQIYIDENRKKAEAHKIFEGLGPKVELPLYNQSDTKVYHENIK 280
Db 403 LPARLPTVDEP--IVISEQHEEDRSATSGAD-----YERSFDQDVTY--K 445
Qy 281 TQVVRKKLILFFKRRNHARKQEQIKQYDQLMEAWE-KKYDRIENPRRKAKESTR 339
Db 446 SSPLLEPSEDVMSKEPELTQEEIDHIAQSLAEKSSFEQSTIPDRPPLPARLPTVD 505
Qy 340 EYEEKQFPIRQORQOQORFVG-----ORGAGLSATIARS-----376
Db 506 E-----PIVSEQHEEDRSATSGADYERSFDQDVTYERSPLLEPCEPVMESKEPEL 559
Qy 377 -RHEISEI-IDGLSQQENNEKQMRQSLVIP-----PMDFDAEORRYKFIN 419
Db 560 TQEEIDNIAQSLAEQSSFE---QASTIPDRPPLPVRLPTVDEPIVSEQOEEEDRSSA 615
Qy 420 MNGL-----MEDPMKVYKDRQFMNVMTDHEKEIFKQKFIQHPKN 458
Db 616 TSGADYERSFDQDVTYERSPLLEPSEDVMSKEPELTQEEIDHIAWI---QSLAEQSS 672
Qy 459 FGLIASYLERKSVPCVLYIYLTKNENYKALVRRYGRGRNQIARPSQBEKV---E 515
Db 673 FEQASTVPDRPPLP-----VRLPTVDEPIVSSE 700
Qy 516 EKEEDKAEKTEKKEEKKDEE-----KDEKEDSKENTKEKDKIDGTAETE 562
Db 701 QHEEDRSATSGADYERSFDQDVTYERSPLLEPCEPVMESKEPELTQEEIDHIAWIOS 760
Qy 563 EREQATPRGRKKTANSQGRKKGR-----ITRSMTNEAASAAAAA-----ATPE 606
Db 761 SAQSSFEQASTVPDRPPLPVRLPTVDEPIVSEQHEEDRSATSGADYERSFDQDQATYE 820
Qy 607 PPPPLPPPEPISTEPVETSR---WTEEME-----VAKKGLVEH-----GR 645
Db 821 RSPLEPCE---EPVMSRVPPELTQEEIDHIAQSLAEKSSFEQSTIPDRPPLPAR 876
Qy 646 NHAATAKMGTKSEACQKNFY---FNYKRRHNL-----NLLQHKOKTSKPKPEERD 695
Db 877 LPTVDEPIVSEQHEEDRSATSGADYERSDQEVYERSPLLEPCEPVMESKEPELT 936
```











```
QY 1312 LKYPQIKRESPIIRAFEGAITKGPYDGIITIKEMGRSIIHEIPRODILTOESRKTPEV 1371
|| || || || || || || || || || || || || || || || || || || || ||
Db 7102 KAPPK-----GPEIS--EKIIPKPKPPKVPKPPAKKVPKPPK-IVVEKVRVPE-- 7151
QY 1372 QSTRPIIEGSIQGTPIKFDNNNGOSAIKHNVKSIIITGSKLSRGMPLLEIVPENIKVVE 1431
|| || || || || || || || || || || || || || || || || || || || ||
Db 7152 ---EPRVP-----PTKPEVLPKPEVVPKVPVP 7178
QY 1432 RGYEDVRAGETVRSRHTSVSSGSPVLRLSTLHEAPKAQLSPGIYDDTSARTPTVSYQNT 1491
|| || || || || || || || || || || || || || || || || || || || ||
Db 7179 PAK-----KPEAPPPKVPKPEAPK-----EVVPEKKVPV----- 7205
QY 1492 MSRGSPMMNRSTSDVTPPNKSTNHRKSTLTPTQRESIPAK---SPVPGVDVPSHSPDP 1549
|| || || || || || || || || || || || || || || || || || || || ||
Db 7206 ---PPKKPE---VPTKVPKPEVKAAPKVPKPEAPKPPKPPKPPKPPKPPKPPKPPK 7257
QY 1550 HHRGSTAGEVYVWSHLPDLPAMPFRLDPAAYLAFQRLSPPTGPGPSQVLYAMENT 1609
|| || || || || || || || || || || || || || || || || || || || ||
Db 7258 PKK-----PEVPPVRVPEV-----KEVVEPKKVPV----- 7283
QY 1610 ROTILNDYITSQOMQVNLRPDVARGLSPREOPLGLPYPATRGCIIDLTNMPPTILVPHPG 1669
|| || || || || || || || || || || || || || || || || || || || ||
Db 7284 ---APKKPEVTPKVP-EAPKEVVPKVP-----VPPPKK 7316
QY 1670 TSTPPMDRITYIPGTQI-----TFPPRPNASMSPCGHPVTHLAAAASAEEREREREK 1722
|| || || || || || || || || || || || || || || || || || || || ||
Db 7317 PEVPP-TKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPK 7366
QY 1723 ERERERIASDLYLRGSEOPGRGSHGYVRSFS-PSVRTQETMLQORPSVFOGTNGT 1781
|| || || || || || || || || || || || || || || || || || || || ||
Db 7367 EVVEPEPAAPQVTPVPPKPNVPKPAVAKKPELPPKVPKVPKVPKVPKVPKVPKVP 7421
QY 1782 SVITPLDPTAQLRIMPLPAGPSISQGLPASRYNTAADALALVALDAASAPQMDYSKTK 1841
|| || || || || || || || || || || || || || || || || || || || ||
Db 7422 PLVPPKKPEP-----PAKVPKVPKVP-----VPEKKVAVPKK 7454
QY 1842 SKHEARLEENLRSAVSEOOQLEOKTLEVEKRSVQCLYTSSAFPSGKPPQSHSVVYS 1901
|| || || || || || || || || || || || || || || || || || || || ||
Db 7455 PEVPPKVPK-----VPKKVPKLEK-----PAVPVPE----- 7481
QY 1902 EAGKDKGPPPKRSYBEELTRGKTTITANFDVITRQIASDKDAREGSSSSSSSL 1961
|| || || || || || || || || || || || || || || || || || || || ||
Db 7482 ---RAESPPPEV-YEE-----PEEIAPEEII-----APEEKVP 7511
QY 1962 SSHRYETSDALEVTSPASSPAPPOEKLQTOYEVVKAQA-----ENDPTROYEGLPHY 2017
|| || || || || || || || || || || || || || || || || || || || ||
Db 7512 PVAEEEEE-----VPPPAVPEPKKIIPKVPKVPKVPKVPKVPKVPKVPKVPKVP 7566
QY 2018 RPQOESPSP-----QQOLPSSQAEGMGQVPRTHRLITL----- 2051
|| || || || || || || || || || || || || || || || || || || || ||
Db 7567 RPPPPPPAPPKEDVREKIFQLKAIPKKVPENQVPKVELTPLKVPGEKKVKRLP 7626
QY 2052 -ADHICQIITQDFARNQVSSQTPQ-----OPPTSIFQNSPSALVSTPVRT 2095
|| || || || || || || || || || || || || || || || || || || || ||
Db 7627 KPEKKEVVLKSVLRKREPEEPEKVPKLEKVKKVPKVPKVPKVPKVPKVPKVPKVPKVP 7682
QY 2096 KTSNRYSPESQAQSVHHORPGSRVSPENLVDSRSGRSPKSPERSHVSSEPYEP-----I 2150
|| || || || || || || || || || || || || || || || || || || || ||
Db 7683 KREKIPEPTKVPKIPKPAIPLPAPEP-----KPKPEAEVKTIKPPVPEPTPIA 7732
QY 2151 SPQVVPVHVEKODSLLLLSQRGAEPAEORNDARSP-GSISYLPSPFTTKLNTSPMVSKK 2209
|| || || || || || || || || || || || || || || || || || || || ||
Db 7733 APVTPVVPVVGK-----AEAKAPKEAAKPKGPIKGVK-----KTPSPTEAER- 7775
QY 2210 QIFIKLNSGGGSDMAAQAQGTETFNLPVTTSGSVSSRGHSFADPASNLGLEDIIRK 2269
|| || || || || || || || || || || || || || || || || || || || ||
Db 7776 ---RKLPGSGGKPPDEA-DFT-YOLKAVPL----- 7802
QY 2270 ALMGSDDKVEDHGVVMSQPMG-----VVPGTANTSVTSGTTRREGDPPSPHSGV 2321
|| || || || || || || || || || || || || || || || || || || || ||
Db 7803 ---KFKVEIKDIIITSEFFVGSSAIFCLVSPSTAITIWMKDGSNIRE-----SP- 7849
QY 2322 CKPKLISNSRK 2334
```

```
Db 7850 -KHREIADGKDRK 7861
|| || || || || || || || || || || || || || || || || || || || ||
RESULT 30
Q9NFS3
ID Q9NFS3 PRELIMINARY; PRT: 16215 AA.
AC Q9NFS3
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE D-TITIN.
GN SLS OR KET OR D-TITIN OR CG1915.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Y.Q., Broadie K.S.;
RT "Characterization of Drosophila D-Titin gene.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271740; CAB93524.1;
DR FlyBase; FBgn010396; sIs.
DR InterPro; IPR001452;
DR InterPro; IPR001777;
DR InterPro; IPR002106;
DR InterPro; IPR003006;
DR InterPro; IPR003598;
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 48.
DR ProSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS00002; SH3; 1.
DR SMART; SM00408; IGC2; 1.
SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 3.5%; Score 442; DB 5; Length 16215;
Best Local Similarity 18.9%; Pred. No. 3.2e-14;
Matches 482; Conservative 363; Mismatches 906; Indels 800; Gaps 117;

QY 10 QCAFSTEQSRYPHSHVQYTFNTRHQQFVAVPDYRSSHLEVSQASQLLQQQ----- 60
|| || || || || || || || || || || || || || || || || || || || ||
Db 5238 QEAFSWRQPREPKQFIQVEDSTLLHLQE-----RHDTQEQLLQQQVPMDRGRK 5287
|| || || || || || || || || || || || || || || || || || || || ||
QY 61 -----QQQLR-----RPSLLSEHP-----GSDRPQERR 86
|| || || || || || || || || || || || || || || || || || || || ||
Db 5288 KPDQPYVQVQQRVKEEFVEKPKTYEEMHDELVEPTIEQPPVPMWGRGKKKPPQ 5347
|| || || || || || || || || || || || || || || || || || || || ||
QY 87 TSYEPFH-----PGSPVDHDSLESKRPRLEQVSDSHFORVS----- 123
|| || || || || || || || || || || || || || || || || || || || ||
Db 5348 KTFEADHDELVEPTVQVQPEVPMWGRGKKKVAQOETVLSQEVVQTSQVVEQIIVEETK 5407
|| || || || || || || || || || || || || || || || || || || || ||
QY 124 -AAVLPLVHPL-PEGLRASADAKDPAFGGKHEAPSSPIS-----GQPCGDDQNA 171
|| || || || || || || || || || || || || || || || || || || || ||
Db 5408 KTAVRVIPPPEPEQKVEQVTLKPTPRPKPEAVKAEELQLKPLRSTRPVQVPEAEOKA 5467
|| || || || || || || || || || || || || || || || || || || || ||
QY 172 ---SPSKSKEELIQSM-----DRVDREIAKVEQOILKLLKKQO-----QLEEEAAKPPPE 220
|| || || || || || || || || || || || || || || || || || || || ||
Db 5468 YEEATDELTEEPVQVQPMWGRGKKKPKQPEVEITPKLEIAVDTLEEEVPEKPTPEQ 5527
|| || || || || || || || || || || || || || || || || || || || ||
QY 221 -----KPVSPPPVEQKHRSIVOLIYD-----ENRKAEEA 250
|| || || || || || || || || || || || || || || || || || || || ||
Db 5528 PQVPLWARGQKKPKQPDQKQELPKSLAIVDTIEDLIKVPQEPQVPLWERRKKKPPQ 5587
|| || || || || || || || || || || || || || || || || || || || ||
QY 251 HKIFGLGPKVELPLYNOPSOTKYHENI-----KTQVMRKLLILFFKRRNHARKQR 303
|| || || || || || || || || || || || || || || || || || || || ||
Db 5588 QDVIE---EKLDV-----APTKEYKAVDVLDPDEPKVEE---KPEPVLW-----QR 5627
|| || || || || || || || || || || || || || || || || || || || ||
QY 304 EQKICQRTDQLMEAEKKVD-----RIENPPRKAKESKTREYEEKQ 345
|| || || || || || || || || || || || || || || || || || || || ||
```



Db 5628 GKXKPKSEPTVEVHPDEVDQAQIETVVKDEMIIVEKRRKIKTKRPKSTKVEELFEBQ 5687  
Qy 346 -----FPEIRKQORQORFORVQORGAGLSATIAARSEHEISEIIDGLSE 389  
Db 5688 PEEIEEPEEVQKEVIEIEIEIEKRLKTKPKLTQQVTEETPHE--EII----- 5740  
Qy 390 QENNEKQMRQLSVIPPMFADABORVKIENMGLMEDPMKYKDRQFMVWTDHKEIKF 449  
Db 5741 KESEVQOEIEIV-----EKKVKVKV-----KPKTVAERQOL-----KEEETPT 5781  
Qy 450 DKFIOHPNFGLIASLYBRKSPDCVLYLYLTKKNENYK-----ALVRNYGKRRGRNQOI 505  
Db 5782 EETVE-----EETRADQOLVVEESKKVKVKKPKTGTVEKTDVDELPGEV-- 5827  
Qy 506 ARPSOEKVEEKEDKAEKTEKEEKEDEEKDEKEDSKENTKEKDKIDGTAEETE--- 562  
Db 5828 --PVEVPVEEVPEVAPE--EELIEEIEVDDEIQEQKRKVKAKKPKTKIETETEI 5884  
Qy 563 -----ERQATPRGRKTKANSQGRKGRITRSMWN-----EAAAAA 599  
Db 5885 EBDQPEEEVLQEEIIEGEEIEITERQKVKSIKKPKKVVTEKTVQDTEQEKPEESQAEEV 5944  
Qy 600 AAAATEEPPLPPPP-----EPISPEVETSWTEEMEVAKKGLVEHGRNWAATAK 652  
Db 5945 KETVTEEPKAPAEBAKVEQEKISLKPAPRKORLLPEKEQVEEVLKPKVKIVAYSE 6004  
Qy 653 M---VGTSEAOCKNFYKRRHNLNLLQHQKTKSRKPREERDVSOCEVASTVSAQ 709  
Db 6005 ABOPETPETEVKEFAIT-----TTEDILDVTKRKKKKPKTK-----VAAE 6048  
Qy 710 EDEDTEAENEENPED--SEVAVKSPEDSNATSRGNTPEAV--ELETTETAPSTPSL 767  
Db 6049 ESTEPAETEFEATEAQPEVQVEEIPPE-----EPQVKEVADERKTA----- 6093  
Qy 768 AVPSKTPAEDESVEVQVNDVSI SAETAQMDVDQOEHSAEEGSCVCDPPPATKADSDVDEVR 827  
Db 6094 --PKPKPKKEELIEKVEEVALKRVTRPKKEL--PQEAITEEVRNL-----KPTQRTSIRKPEV 6147  
Qy 828 VPE---NHASKVEGONTKERDIDRASEKV--EPRDSDLVVAQOINNAORPEQSDNDSSATC 883  
Db 6148 LEEVDLQHVKEKEDIQVEE--KRTRKVKPKKHEDL-----PEIPDAEPQL 6193  
Qy 884 SADEVGDGEPRQRMPMDSPKSLNPTGSLVSLSPKPNPLD-----LP 928  
Db 6194 BEAEHIELEKQPK---PEEDQPVMPKRG-----KKQPVVEEVEEKWPSGKRRRLP 6243  
Qy 929 QLOHRAAIVPMVSWCTPCNIGTPVSGYALYQRIKAMHESALLEEORQOEIDLECR 988  
Db 6244 BQPEEVQKPKPS-----KPIEE---QOKPEKAIPQPLVPEKPESEEELE-- 6289  
Qy 989 SSTSPCGTSKSPNREWEVLOPAPHQLITNLPEGVRLPTTRTPRPPLIPSSKTTVASEK 1048  
Db 6290 -----LEP-----LKLPEDK--KPEKPAKKEK-----KKK 6313  
Qy 1049 PSFINGGISOGTGTLYTSHNQA-----SYTOETPKPSVSGISLG---LPROQE 1095  
Db 6314 P-----KLKATPSVDEVEEVAEPDEPTAEDEVEEMPVDDVKVAVSEDLVLP-BEE 6366  
Qy 1096 SAKSATLPIKQEEPSRSONSQPGLLVRAHQHGVGRTAGATQEGSITRGTPTSKISV 1155  
Db 6367 VVPTETPEAKOKAHKRTKR-----LKEASVGEQOLLEAAIAE-----I 6407  
Qy 1156 ESIPSLRGSITQGNPAL-----PQGTPEALVKSGIS----- 1188  
Db 6408 EKVDEISEIJSQKTLTLKKTEDTRPQF--ITEQLIEDVDEVRDLKMKVTSNIIEK 6466  
Qy 1189 -RMPEDSSPEKGREAAKGHVIEY-----KSGHILSYDNKINAREGTRSPRTAHEIS 1242  
Db 6467 RRVVLDDSOPLLELITQKR--IQEGDKVADEELIEDQQLIQNQOETTTTSEVIGQERK 6524  
Qy 1243 L--KRSYVESVGNKQGMRESVSPAPLEGICRALPRGSPHDLKERTVLSGIMOGT 1300  
Db 6525 LVKKKKKIKPPIREKLRPRQ---CVPEEPTVLCKVEGVFPFPEIK----- 6568

Qy 1301 PRATTESFEDGL-----KYPKOIKRESPPIRAFEGAITKGPYD--GIITIKEMGRSIIHEI 1354  
Db 6569 -----WYFNIDILLFASEKYEIVTMEQVAKL-----IAKVTSDGVGYTCEAKNEAGVAT 6618  
Qy 1355 PQODILTQESRKTPEVOSTRPIIEGSIQSGTPIKFDNNNGSQAIAKHNVKSLITG--PS--- 1411  
Db 6619 SRTNIIILEEQGVPP--QFTKPL-----KIBFIEEKQPERLKVTVCQVTKGPNPE 6667  
Qy 1412 -KLSRGMPLEIVP--ENIKVVERGKVEDVKAGETVRSRHTSVVSSGSPSVLRSTLHEAPKA 1469  
Db 6668 VKWYKGE--EVIPSETVOM-----FYDEKTDGVAL-----EVINPTPNEAV-- 6707  
Qy 1470 QLSPGIYDDTSARRTPVSYQNTMSRSGPMNRTSDVTIPPKNSTNHERKSTLTPTQRESI 1529  
Db 6708 ----VYS-----VQAQNGFRAIGNANILSRVDEVPREIL---KAPTVTPLSAVVV 6751  
Qy 1530 PAKSPVPGVDVPVSHSPDPHHRGSTAGEVYVWSHLPTQLDPAHPHRAALDPAAYLFOR 1589  
Db 6752 PTGGTL-----FFEAKYDGLPRPEVKWRNGREI---IENEETIETTTTITIK 6798  
Qy 1590 QLSPTPGYPSOXYLA-----MENTROTILNDYITSOOMOVNLRPDVARGLSPR----- 1638  
Db 6799 VVNMTRKRTGKTEVWAKNKVGEAKSSGVVSDQKPDQIK---PPRFTQPLEPKYFGEH 6855  
Qy 1639 -----EQPLGL-----PYPATR--GIIDLTMNPPTILVPHPGTSTTPMDRI 1678  
Db 6856 EVAIEAIEVESEPLSOFQWVNEPIKSSNEVRIYSQAN--KSTLLIENQSKFVGP--- 6910  
Qy 1679 TVIPGTQITFPFRPNYSASMPGHPHTLAAAASARERERERERERERERERERERERERER 1738  
Db 6911 -----FTCRAENVG---GSVTSTATNLIPOEAEFEESPFRVEEL-----V 6949  
Qy 1739 RPSGEPQGR-----PGSHGVRSPPSV---RTOETMLQORPSVFOGTNGTSTVITPL 1787  
Db 6950 QPVEVMDGEALLTLCQVGT-----KPTPKVEYHNAEKITENKE-----TTISQDL 6995  
Qy 1788 DPTAQLRIMPLPAGGPSISQGLPASRYTAADALAALVDAASAPQMDVSKTESKHEAA 1847  
Db 6996 QGVCOQLQITEVPEPENGQYECVATNKGKSVKTVNVIQAFYIPDSEITGLTS----- 7050  
Qy 1848 RLEENLRSSAAVSB0-----QOLEQKTLVEKRSVOCLYTSSAPPSPKQPQ--- 1894  
Db 7051 --EEDLLDRTLSIDQAPKIIKKLPEKIEPKEGEQAKLEVWV-----GKPKPKVKW 7100  
Qy 1895 -----HSSVVSAGKDKGPPP---KSRYEEELTRGKTTITTAANFDIVIT----- 1938  
Db 7101 LRDDQIIFASEEYQIENFEDGTSVLVINHVYPDDL---GTISFEAYNPLGVAVTTALFAV 7157  
Qy 1939 RQIASDKDAR--ERGSQSDSSSLSSHRYETPDAIEV-----ISPASP 1982  
Db 7158 EGIVGSKDYRKPEWTSQMEQVAKAAKC--SPSLLNEMDRCAALGETAKESIQFAGNP 7216  
Qy 1983 APPQEKLOTYQEV--VKANAQENDPTROYEPLHHYRFQOESPSPQOQLPPSSQAEQMGQ 2041  
Db 7217 IP---DIQWYFNWQLRASEKYRVWVYQOEAATLTK----- 7250  
Qy 2042 VPRTHRLITLAD---HICQII-----TODFARNVYSQTPQOQPTSTFONSPSALVSTPV 2093  
Db 7251 -----ITSEDCGYNCKLINEIGMTMTAKFDISS-----TSTIVEETKAKTVKK 7296  
Qy 2094 RT--KTSNRYSPESQAQSVHHORPGSRVSPENLVKSRGSRPKSPERSHVSSEPEYEPIS 2151  
Db 7297 KSGKTKWKRSGASSQNV--QKTEIRIIPITSAVTSMNVKVKQP----- 7340  
Qy 2152 PPQVPVVEHEKOD--SLLLLSQGAEPAEQRN 2180  
Db 7341 ---VSVLVEKSEISEVLVVKDREVADEERS 7368  
RESULT 31  
Q17490  
ID Q17490 PRELIMINARY; PRT; 5170 AA.



AC Q17490;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE CODED FOR BY C. ELEGANS CDNA YK11B8.3.  
GN B0350.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Gattung S.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U50071; AA93447.1;  
SQ SEQUENCE 5170 AA; 575856 MW; 27C2DFBF4AE03A7B CRC64;

Query Match 3.5%; Score 440; DB 5; Length 5170;  
Best Local Similarity 18.7%; Pred. No. 8.7e-15;  
Matches 556; Conservative 357; Mismatches 1101; Indels 958; Gaps 126;

QY 14 STEQRYPPHVSQVYTFNTHQOEF-----AVPYRSHLEVSQASQLQOQ 60  
DB 775 SSEQSEPHIVKETTITVTRELYDEPEKGNVTFSPAPSSHAESRQVPSVSHQE 834  
QY 61 QOQ--OLRRPSSLSEFH-----PGSRP-----QERTSYEPHP-GPSVPHDSLES 106  
DB 835 YPHVVETTTTNTNIYDDEDNVPSSEDPATQHFQOQSETSVHRSHPDVSESDGGLGS 894  
QY 107 KRPRLEQVSDSHFQVSAAVPLVHPLPEGLRASADA-KKDPAFGKGKEAPSSPI-----S 161  
DB 895 KVLGFAKKAGVAGVGAAPVALA---AVGAKAAYDAFEKDEDDDETSHSPSPVPEQOS 951  
QY 162 GPCGDDQNAS-----PSKLSKEELIOSMRVDREIAKVEQOILKKKKOQOLEE 212  
DB 952 EOYQDQASQSHDTFEHMPESPHEHEKETEFDH-----SHPESPVLSKEREHQVTS 1006  
QY 213 AA-----KPEPEK-----PVSPPP-----VEQKHSIVQII----- 239  
DB 1007 TTTTITVTREYNDEPQEHQGHPSAPSSHTAEHPHIVETTTTITVTREFQPEPEL 1066  
QY 240 ---YDENKKAEAKHIFEGLGPKVLPYLNQPSDTKYVHENIKTNQVMRKLILFFKRR 296  
DB 1067 EYKQEDNRKSPSSH-----SQENLVTTTTTITVSE-----YDEP 1104  
QY 297 NHARKQEQIKQYDQLMWEAKV--DRIENPRKAKESKREYKQFPEIRKQRE 354  
DB 1105 EHFEQETGKSPAPSSHVES-ERQVLESPVASDPDH-VMTITTTTITVTROFHDDESERS 1162  
QY 355 QOERFORQORCAGALSATIASEHSELIDG-----LSEQE--NNEQMRQLSVIP 404  
DB 1163 DSPNRDVV--EAQSIHSSNIHQOFSEKDDDSQGRGSPVKSEDEPVKHESYKFTETSTTE 1220

QY 405 PMMFDAEQRRVKFINNGLME---DPMKVYKDRQPMNVMTDHEKEIFKDKFIOHPKNE-- 459  
DB 1221 TREFDRPE-----LDSERLSEPAQSPMEV-----SPITTEENIVRSESSVK 1261  
QY 460 -----GLIASYLERKSVPCVLYYYLTKKNENYKALVRRNYGKRRNQOIARPSQEEK 513  
DB 1262 SEYGSEGHVPVSVIETKTTTITVTREFYDDQDE-----HENQTOSEELRSSIPTEEEE 1313  
QY 514 VEK---BEDKAETEKKE--BEKDEEKEDEKEDSKENTKEKDKIDGTAETETEREQAT 568  
DB 1314 HEGSHLFKETTITVTREFYDEPENVELOPQSPAPSSHVESGIIHASESPVAAQOEI 1373  
QY 569 PRGRKTSQGRKGRITRMTNEAAA-----SAAAAATEEPPLPPPEPISPEP 622  
DB 1374 PQ-----TREFHEDSPAAYFHEEYESHVLTEQAAPLLTEQHOQSPESGDE 1418  
QY 623 VETSRWTEEMEVAKK-GLVEHGRNMAAIA-KMVGTKSEAOCKNFYNYKRRHNDNLQ 680  
DB 1419 SDGEGLSKVLGFAKKAGVAGVVAAPVALAAVCAKAAKDALKKDDDEDEQERESLLR 1478  
QY 681 QHQ-----KTSRPRERDYVQCESVASTVSAQDE-----DIEA 716  
DB 1479 QERSIDSPHASEQOIEBEHERFEESPSEKHVHTETTTTITVTREFYDEHEPLVSOEI 1538  
QY 717 SNEENPEDSE-----VEAVKPSDSDENAT-SRGNTEPAVE 752  
DB 1539 EGEKSPADSEKSLPHVVETTTITVTREFDKNDSFVPSPEKEQPTVSEVYETABG 1598  
QY 753 LEPT--TETAPST-----SPSLAVPSTKPA-----EDESVE 781  
DB 1599 DEPEHYTETTTTITVTKEVIDDSQMGDDPEKQSPQVETTTTITMTSREYDDNDETR 1658  
QY 782 TVNDISIAETAQMDV-----DQEHSAEGSVCDPPP-----ATKADSDV 824  
DB 1659 SEAGSHITETKTITTVVREFHGEQPEETETDEVEELPPKIEEDNVSESESTSVSR 1718  
QY 825 EYRVPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQIQAQRPQSDSSATCS 884  
DB 1719 EVRPDEPHIETTTTITVTREYHNEPEETDYDQKD--AAPISFS-QEHQDDDSQAG-- 1771  
QY 885 ADEDVGGPEQRMPMDKSPKSLNPTGSLVSSPLK-----PNPLDLPOLQH 932  
DB 1772 -----HDQHDRESFVSEKSVKHTTETTTTITVTROLYDDDEASEIRGESPVATEEH 1824  
QY 933 RAAVTPPVSTPCNIPIGTPVSGVALYQORHIKAMHESALL-----EEOQRORQ 982  
DB 1825 -----VSSTKSDDE-----SEQHPVSVIETTTTITVTREFYDDQDELQREDH 1866  
QY 983 IDLECRSTSPCGTSKSPNREWEVLQAPAPHOLITNLPGVRLPTTRPTR---PPPPLIPS 1039  
DB 1867 TQSEERRSIP--TEETEHEDSHLIKET-----TTTTTITVREFYDEPENVEK 1911  
QY 1040 SKTTVASKEPFSIMGGSISQGTPTGYLTHNQAS-YTOETPKPSVGSISLGLPQROESAK 1098  
DB 1912 LQDSQFSLSPS-----SHVESEIYVPEP-----VAKQOEIPQ 1944  
QY 1099 -----SATLPYIKQEEFSPRSONSQOEGLLVRAQHEGVVGTAGATOESITRGTPS 1151  
DB 1945 TREFHEDSPAAYFHEDEYEHQVPTQAP-LLTEQOHP-----ESGEESGEGFG 1994  
QY 1152 KISVESIPSLRGSTQGTPTALPQTGIPTEAL-VKGSISRMPIEDSSPEKREPEAAKSHV 1210  
DB 1995 S-KVLGFAKKAGVAGVVAAP---VALAAVCAKAAKDALKKEDD--BEDQERES---L 2045  
QY 1211 IYEGKSGHILSYDNITKNAREGTSRPTAHEISLKRSYSEVGNITKQGMRESPPVABLE 1270  
DB 2046 IREER-----SFD-----SPHASEQOIEKEHRFE-----ESPVPSEKH 2080  
QY 1271 GLICRALPRGSPHSDLKE--RTVLGSGSIMQGTPTTRATTSFEDGLKYPKOIKRESPIRAF 1328  
DB 2081 HDQSALPQESVSPQIEKESRTFNDESEFGVKSDDHYEDDQESLKSPE----- 2129



QY 1329 EGATKGPYDGIITIKEMGRS-----IHEIPRODILTQSRKTPVQSTRPIIE----- 1379  
Db 2130 -----SGEAFSQFTSEKQDRSDPIHS-QKEDISQFQNESQEDVKSEQHDEKPDLE 2183  
QY 1380 --GTSQCTPIKFDNNQSQAIAKHNVKSLING--PSKLSRGMPLEIIVPENI----- 1427  
Db 2184 RQGSYSSGYSKSPGGS-----ITGLDEKALSQVQPEPDRPFAESHEKTE 2231  
QY 1428 -----KVVERGKYEDVAKAGETVRSRHTS--VVSSGSPVLRSLTHEAPKAQLSPGIYDQTS 1480  
Db 2232 ATSDENLFESDKY---APASVPSESDSNRVIEITTTTITVREHFEPE-----DHS 2280  
QY 1481 ARRTPVSTQNTMSRGSPMMNRSTSDVTPPNKSTNHERKSTLTPT-QRESIPAKSPVPGVD 1539  
Db 2281 ---YVSESQYSSSGSP-----VPSEKSDVRVIEITTTTITVREHFEAEDEIPI- 2327  
QY 1540 PVSHPSPDPHRRSTAGEVYWSHLPTQLDPAMPFHRALDPAALAYLQFOLSTPTGPVS 1599  
Db 2328 -----VSSH-----DDQAA-----SPVSEED 2345  
QY 1600 QYQYAMENTROTIL-----NDYITSOOMQVNLRPDVARGLSPREQLPLGPLYPATRGII 1653  
Db 2346 VHEQIQTITTTTITVTKHEFVDDDEIDSEHM-----ESDKYASGSPVSEEDS 2394  
QY 1654 DLTNMPPILVP-----HPGGTSTPMDRITYPGTQITFPFPPYNSASMSPGHPTHLAAA 1709  
Db 2395 RVETITTTTITVREHFEDEDDHSPVVOQY-----SASESP-----VPSE 2435  
QY 1710 ASAR-----ERERERERERERERIAAASDLYLRPGSEQP-- 1745  
Db 2436 KSERVIEITTTTITVREHFEDEHILGQGESDSQIPSESITSENMD---RETSSSPVQ 2492  
QY 1746 -----GRGSGHYVRSPSPSVRT-----QETMLQORPSVFOCTN 1779  
Db 2493 SNRDEEFVLPALPYKQTEYGRVDSDHAPASAPASBPSTQAYKQESQEAHLEQFQ 2552  
QY 1780 GTSVITPLDPTAQL-----RIMPLPAG-GPSISQ----- 1807  
Db 2553 RSVSHEDSPAAQYFHONESDHDSPVSDRAPLLTQOQHQPESGSDGEGFGKVLGFA 2612  
QY 1808 ---GLPASRYNTAADALAALVDAASAPOMDVSKTESKHEAARLEENLRSRAVSEQ 1864  
Db 2613 KKAGVAGVVAAPVALAAGAKAA---YDALKKDDDEDPENDEQEKLLPKS---PERQ 2665  
QY 1865 QL-----EQKLEVEKRSVQCLYTSTSAFSPGKPOPHSSVYSEACKGKPPPKSRVEELR 1920  
Db 2666 VLVNVPESDSEISE--ILEYTS-----PSTKSESQCY 2699  
QY 1921 TRGKTITTAANFIDVITRQIASDKDAREGSSQSSDSSSLSSHRYETPDAIEVSPAS 1980  
Db 2700 TETVTRTT-----VTREYLD-----POSVTRSRSPSEH-----DISEQYAPES 2738  
QY 1981 -----SPAPQE-KLOYQYQEVVYKANAENDPTQYEGPLHY--- 2017  
Db 2739 PVEQDPYVVEKTTVIRQYHDEPQOEIEEQTIPEVTVLREVYESP--EGDEPEOHYET 2796  
QY 2018 -----VTREYLD-----RPOQESPS----- 2025  
Db 2797 KTTTTITKEVHPVEDVQISPVHSETSVSEKQLPADEQLDEPVTETTTATVTRERYE 2856  
QY 2026 PQQLPSPS-SQAEQMGQVPRTHRLITLADHICQI-ITQDFARNQVSSGTPOQPTSTFTON 2083  
Db 2857 PEEVPPSPGSEADDESHAPKYMETTTTTTITVREYEVSEDEHQQSQVQORDSPAPSEDS 2916  
QY 2084 SPSALVSTPVTKTSNRYSPESQASV-----HHOPCSRVSPE 2122  
Db 2917 VKHIEKTTTTTITVTEPEDEDSHPVSEDDVHGFKVTTTTTITVTHEPEPEPPSPDE 2976  
QY 2123 NLVDKSRGSRPKSPERSHVSSEPEVPISSPQVPVWH-EKODSLLLLSQCAEPAEQRND 2181  
Db 2977 HVVESERYASGSPVSEED--SSRIETTTTTTITVREHFELED-----DOEHVESQEYSA 3030  
QY 2182 ARSPGSIYLPSS-----FFTCLNTSPMVYKSKQEIFRKLNSGGGSDMAAAQFCTE 2234

Query Match

3.5%; Score 438; DB 5; Length 6994;

Db 3031 SGSP-----VPSEKSERVIEITTTTITVREHFEHDDIPTIVTSHDDPAASSVPSEED 3085  
QY 2235 IFNLPAVTTSGSVSRGSHFADPASNLGLEDIIRKALMGSPDDKVEDHGVMQSPMGVVP 2294  
Db 3086 VHGOIQTTTTTTTTTITVREHVPVDEEIDSGRMDLEKYSSESPVSEEDSSRVIE----- 3138  
QY 2295 GTANTSVTSGTREGDPSPHSGVCKPKLISKNSRKSKSPIPGQ----- 2342  
Db 3139 -TTTTTIVS-----REHFEPEDDHSHVSESQFYSASG-----SPVSEKSERVIEITTTT 3187  
QY 2343 -----GYLGTER-----PSSVSVSHSEG--DYH--RQTPGWAMEDR----- 2374  
Db 3188 TTVTREHFDTEEDYIPSESRTHDDGITDQHVPSQSPVSEEDDHVTHEQIIRDDPIDEQ 3247  
QY 2375 -----PSSTGCTOPFPNPLTRMLSLSTPTPTPIACAPSAVQAAPHQONRI-- 2419  
Db 3248 IVESHEYKSESIPEQDSTH-----VIETVMTSPITTSERYDPEVEKDVVESADDEIDS 3301  
QY 2420 -----WEREPAPLLSAQYETLSDD 2440  
Db 3302 STAQYKSESPVQTEKSLLLAKQQQESGDES 3333

## RESULT 32

Q17343  
ID Q17343 PRELIMINARY; PRT; 6994 AA.  
AC Q17343;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE UNC-44 ANKYRINS.  
GN UNC-44.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2;  
RX MEDLINE=95263663; PubMed=7744957;  
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,  
RA Boontrakulpoontawe P., Jeyaprakash A., Hedgecock E., Wheaton V.I.,  
RA Sobery A.;  
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal  
RT guidance in Caenorhabditis elegans.";  
RL J. Cell Biol. 129:1081-1092(1995).  
RN [2]  
RN REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.  
RC STRAIN=N2;  
RA Otsuka A.J.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U39847; AAB41827.1; -;  
DR EMBL; U21733; AAB38384.1; -;  
DR HSP; P42773; IHHB.  
DR InterPro; IPR000488; -;  
DR InterPro; IPR000906; -;  
DR InterPro; IPR001360; -;  
DR InterPro; IPR002110; -;  
DR InterPro; IPR002383; -;  
DR Pfam; PF00023; ank; 24.  
DR Pfam; PF00531; death; 1.  
DR Pfam; PF00791; ZU5; 1.  
DR PRINTS; PR00001; GLABLOOD.  
DR PROSITE; PS00088; ANK\_REPEAT; 22.  
DR PROSITE; PS00297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00017; DEATH\_DOMAIN; 1.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; UNKNOWN\_2.  
DR SMART; SM00005; DEATH; 1.  
SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;



Best Local Similarity 18.7%; Pred. No. 1.7e-14;  
Matches 556; Conservative 356; Mismatches 1102; Indels 958; Gaps 126;

```
QY 14 STEQSRYPHPSVQYTFPNTRHQEE-----AVDPYRSSHLEVSQASOLLQOQ 60
Db 2599 SSEQSEEPHIVKETTITTTTTRRELYDPEKGNVTFSPAPSSHAESRQVPSPVVSQHE 2658
QY 61 QOQ--QLRRRPSLLSEFH-----PGSRP-----QERTSYEPFHP--GPSFVDRHDSLES 106
Db 2659 YPHVVETITNTVTSNIYDDENVPSSDPATQHPQOSETSVHRSHPDPSVEESDGEGLS 2718
QY 107 KRPRLEQVSDSHFORVSAVPLVHPLPEGLRASADA--KQDPAFGKGHEAPSSPI-----S 161
Db 2719 KYLGFPAKAGVAGVAAVALA---AVGAKAAYDAFEKDEDDDETSHSPESPVPYQS 2775
QY 162 GQPCGDDQNAS-----PSKLSKELIQSMRDVREIAKVEOOILKLKKQOOLBEE 212
Db 2776 EYQDDSAQSSHTDEHHMPEPIHEKETEEFDH-----SHPESPVLSEKERDDQVTSSE 2830
QY 213 AA-----KPEPEK-----PVSPPP-----VEOKHRSIQIIL-----239
Db 2831 TTTTITTTVREYNDEPEDQEHQHPSPAPSSHTAEAPHIVETITTTTITVTRFQEEPEEL 2890
QY 240 ---YDENRKKAEAAHKEFGLGPKVELPLYNOPSDTKYVHENIKTNQVMRKLILFFKRR 296
Db 2891 EYQEDNRSKSPSSH-----SOENLVETITTTTITVTS-----YYDEP 2928
QY 297 NHARKQEQKICQRYDQLEMAWEKKV--DRIENPRRAKESKREYYEYKOPPELIRKQRE 354
Db 2929 EHFEQETGKSPAPSSHVES--EROVLESPVASDPRH--VMETITTTTITVTRFHDDESRS 2986
QY 355 QOERQVRQGRGAGLSATIAARSEHSEISIIDG-----LSEQE--NNEKQMROLSVIP 404
Db 2987 DSPNRDVV--EAQSIHSSNIESHQOFSEKDDDSQGRGSPVSEDEPVKHEGKYKFTETSTE 3044
QY 405 PMMFDAEQRRVKFINNGLME---DPMKVYKDRQPMNVMTDHEKEIFKDKFIQHPKNF--459
Db 3045 TREFDRPE-----LDSERLSEPAQSPMEV-----SPITTEENIVRPSSSVK 3085
QY 460 -----GLIASYLERKSPYDCVLYLYLKKNENYKALYRNYGKRGRNQOIARPSQREK 513
Db 3086 SEYSGEHPVSVIEKTTTITVTRFYDDQDE-----HENQIQSEELSSSIPTSEEE 3137
QY 514 VEK---BEDKAEATEKKE--BEKKDEEEKEDKESKENTREKDKIDGTABETEEREQAT 568
Db 3138 HEGSHLFKETTITTTTITVTRFYDEPENVEELQDPQSPAPSSHVESGIIHASESPVAQOEI 3197
QY 569 PRGRKTANSQGRGRGRITRSMNTNEAAA-----SAAAAATEEPPLPPPPFPFISTEP 622
Db 3198 PQ-----TREFHSDPAAQYFHEEYESHVLTEQAPLLTEQOQHQPESGDE 3242
QY 623 VETSRWTEEMEVAKK--GLVEHGRNWAIA--KMVCTKSEAQCKNFYFNKRRHNLNLLQ 680
Db 3243 SDGELGSKVLGFAKAGVAGVAAVALA--AAVAAKAAAYDALKKDDDEEDQEERESLLR 3302
QY 681 QHKQ-----KTSRPREEDVYSCQESVASTVSAQED-----DIEA 716
Db 3303 QERSIDSPHASEQSQIEEHEHFEESPVPSEKHVHTETITTTTITVTRFYFDEHEPLVSQEI 3362
QY 717 SNEENPDSE-----VEAVKPSDESPENAT--SRGNTEPAVE 752
Db 3363 ECKGSPADSKSLPHVVVETITTTTITVTRFEDKNDSESPVSEKEQEPITVSREVYETAE 3422
QY 753 LEPT---TETAPST-----SPLAVPSTKPA-----EDESVE 781
Db 3423 DEPEHHYTETITTTTITVKEVIDDSQEMGDDDEPKQESQOVVETITTTTITVTRSYDDNDDETR 3482
QY 782 TVQVNDISAEATAEQMDV-----DQOEHSAEGSVCDPPP-----ATKADSYDV 824
Db 3483 SEAGDSHITETTKTTTVVREHFGQEPETEETDEVEELPPKIEEDNVDNVEYSESSSTSVSR 3542
QY 825 EYRVPENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQINAOIRPEPQSDNDSSATCS 884
Db 3542 EYRVPENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQINAOIRPEPQSDNDSSATCS 884
```











DR	InterPro: IPR003006; ..	3 4%;	Score 432.5;	DB 15;	Length 3263;
SR	Pfam: PF00047; Ig: 7.	Best Local Similarity 18.7%;	Pred: No. 1.2e-14;		
DR	SEQUENCE 3263 AA; 373083 MW; 3776DE5CD75E03F3 CRC64;	Matches 472; Conservative 361; Mismatches 905; Indels 791; Gaps			
QY	10	QGAFSTEQSRYPHSHVQYTFPNTRHQQEFAVPDYRSSHLEVSQASQLLQQQ	-----	60	
Db	473	QEAFSWQRPQKFTIQVEDSTLLHLQE	-----RHDTQEQLLQQQPVVMWDRGRK	522	
QY	61	-----QQOQLR-----RPSLLSEPH	-----	86	
Db	523	KPDQPOYVQPOQRVKKEEFVKPTYEENHDELVEPTIEQPPVPMWGRGKKKPPQ	-----	582	
QY	87	TSYEPEH-----PGSPVDHDSLESKRPLEOVSDSHFORVS	-----	123	
Db	583	KTEEAHDELVEPTVQOPEPVPVPMWGRGKKKVAQOEIVLSQEVVQTSQVVEEQIVETK	642		
QY	124	-AAVLPLVHPL-PEGLRASADAKDPAGFGKHEAPSPIS	-----GQPCDDQNA	171	
Db	643	KTAVRRVIPPPEPQVEQVTLKTPRPRPKAEVAAEETQLKPLRSTRPVPQVPEAEQA	702		
QY	172	---SPSKSKEELIQSM---DRVDREIAKVEQOILKLKKKQ	-----QLEEAAKPPEPE	220	
Db	703	YEEATDELTEEP1PQPPVPMWGRGKKKPKQPOEEVTEPKTLEIAVDTLEEEVPKPTPEQ	762		
QY	221	-----KPVSPPPVEQKHSRVIQIYD-----	ENRKKABEA	250	
Db	763	PQVPLWARGKKPKPQDEQKQLPKSLEIAVDTIEEDLIKVPQEPQVPLWERKKKPKP	822		
QY	251	HKTFEGLGPKVELPLYNQPSDKVYHENI-----KTQVMRKKLILFFKRRNHARKQR	303		
Db	823	QDVIE---EKLDV---APTKTYEKADVLPDPKVEE---KPEPLW	-----QR	862	
QY	304	EOKICORYDQLEAWKKVD-----RIENPRKAKESKTREYKEK	345		
Db	863	GKKKIPKSPTEVHPDVEDAQTIVVKDEDMIVBEKRIKTKRPKSTKEVTEELFEEQ	922		
QY	346	-----FPPIRQOREQERFORVGQRGAGLSATARSHEISEIIDGLSE	389		
Db	923	PEEISPEEVPQKEVIEIEEIVEEKRLKTKKPKLTQQVTEETPEH-EII	-----975		
QY	390	QENNEKQMQLSVIPPMPFDAQRRVKFNNGLMEDPMKVYKDRQFMVNTDHEKEITFK	449		
Db	976	KESEEVVQEEIV-----EBKKKKVKV-----KPTVAEKQL-----KEEITPT	1016		
QY	450	DKFIQHPKNGFLIASLYERKSPYDCVLYLYLTKNENYK---	ALVRRNYGKRGRNQI	505	
Db	1017	EEIVE-----EEETAEDQLVVEESKKVKVKKPTGTVKEKTVDELPGEEV	-----1062		
QY	506	ARPSQEEKVEEKEDKAERTEKKEEEDKDEKDEKDESKENTKEKDKIDGTABETE	---562		
Db	1063	--PVEEVPVEEVPEDVAPE-EELIEEQQEIVDQEIQGRKKVKKAKPKTKIEKTEI	1119		
QY	563	-----ERQATPRGRKKTANSGRKGRGRTSMTN-----EAAAASAA	599		
Db	1120	EEDQPEEVLQEEIIGQEEITERQKVKSIKKPKKVVYETVDTQTEQPKPEESQAEV	1179		
QY	600	AAAAATEPPPLPPPP-----EP1STPEVTSWTTEEMEVAKKGLVEHGRNWAATAK	652		
Db	1180	KEVTVEPKPKPAPEAAKVEQVEKISLKPAPRKQRLLPKEQVEEVLKPKVKIVAVSE	1239		
QY	653	M---VGTKSEAQCKNFYFNKRRHNLNDLNLQHQKQKTSRKPREERDVSCQESVASTSAQ	709		
Db	1240	AEQPTPETEFVKEFAIT-----TTEDILDVTKRKKKKPKTK-----VAEE	1283		
QY	710	EDEDIASCNEENPDP-SVEAVKSESDESPENATSRGWTPEAV-ELEPTTTAPSTSPSL	767		
Db	1284	ESTPEAPEETEFEAEATQPEEVQVEETPE-----EPQKVEADERTK-----	1328		

Qy	768	AVPSTKPAED	SVETQVND	SI	SAET	AOQMDV	DQOQ	HS	AE	GS	VC	DD	PP	PA	T	K	A	D	S	V	D	V	E	R	827	
Db	1329	--PKP	PR	KE	II	TE	K	V	E	A	L	K	R	V	T	P	K	K	E	L	--P	Q	E	A	T	138
Qy	828	VPE	--NH	ASK	Y	E	G	D	N	T	K	E	R	D	L	R	A	S	E	K	V	--E	P	R	D	883
Db	1383	LEE	V	D	L	O	H	V	E	K	K	E	I	V	O	E	--K	R	T	R	K	V	K	K	H	1428
Qy	884	S	A	D	E	D	V	G	P	E	R	O	R	M	P	D	S	K	L	N	P	T	S	I	L	928
Db	1429	E	E	A	H	I	E	L	E	K	O	P	--P	E	O	Q	O	V	P	W	K	R	G	E	--K	1478
Qy	929	Q	L	O	H	R	A	A	V	I	P	P	M	V	S	C	T	P	C	N	I	P	I	G	T	988
Db	1479	E	O	O	P	E	E	V	O	L	K	I	P	S	-----K	P	I	E	--O	O	K	E	K	A	I	1524
Qy	989	S	S	T	S	P	C	G	T	S	K	S	P	N	R	E	V	I	L	Q	A	P	H	L	I	1048
Db	1525	-----L	E	P	-----K	L	K	P	E	K	--K	P	O	V	E	E	V	L	E	K	K	W	P	S	K	1478
Qy	1049	P	S	F	I	M	G	S	I	S	O	G	T	P	G	T	L	S	H	N	O	A	-----S	V	T	1095
Db	1549	P	-----K	L	K	A	T	P	S	D	E	V	E	E	V	A	E	P	F	E	D	I	A	E	E	1601
Qy	1096	S	A	K	S	A	I	L	P	I	K	O	E	F	S	P	R	O	N	S	O	P	E	L	L	1155
Db	1602	V	P	T	E	E	T	P	E	A	K	O	A	H	K	R	T	-----L	K	E						



```
Db 2185 QPVEVMDGEALLLTCQVTG-----KPPKVEWYHNAEKITENKE-----TTISQDL 2230
Qy 1798 DPTAQLRIMPLPAGPSPISQGLPASRYNTADALALVDAASAPQMDVSKTKESKHEAA 1847
Db 2231 QGVGCLQITEFPPENEGQYECVATNKIGKSVKTVKIQAFEYIPDSEITGLTGS----- 2285
Qy 1848 RLEENLRSRSAVRSQ-----QQLEOKTLEVEKRSVQCLYTSSAFPSPKQDP----- 1894
Db 2286 --EEDLLDRTLISIDQAPKIITKLPKEIEPKEGEQAKLEVKV-----GKPKPKVKW 2335
Qy 1895 -----HSSVYISAGKDKGPPP---KSRVEEELRTGKTTITAAANFIDVLIITQIASD 1944
Db 2336 LRDDQIFASEYQIENFEDGTSVLVINHVYPDDL---GTISFEAYNPLGVAVITLFA- 2391
Qy 1945 KDAREGQSDDSSSLSHRYETPSDAIEVISPASSAPQOEKLOTQOPEV-VKANQAE 2003
Db 2392 -----VEGIVGSKDYRKD-----EWVSQEMEQQVALKDQIOWTFNNVQLRASEKY 2435
Qy 2004 NDPTROYEGPLHHYRPQESPSQOQLPPSSQACMGQVPRTHRLITLAD---HICQI- 2059
Db 2436 RMVQEQEATLEIMK-----ITSEDCGYYNCKLIN 2465
Qy 2060 ----TQDFARNQVSSOTPOQPTSTFONSPSALVSTPVRT--KTSNRYSPESQASQVHHQ 2113
Db 2466 EIGMTMTAKFDISS-----TSTIVETRAKTTVKKSGKKTVMKSGASESNV--Q 2516
Qy 2114 RPSRVSPENLVDRSGRSPKSPRSVSSEPEYPIPPQVPVHVHKQD--SLLLLSQR 2171
Db 2517 KTEIRIIPTSAVETSMNVIKVKQP-----VSVLVEKSEISeVLVWVDR 2559
Qy 2172 GAEPAEORN 2180
Db 2560 EVADAERS 2568

RESULT 35
Q9VSK5
ID Q9VSK5 PRELIMINARY; PR7; 1963 AA.
AC Q9VSK5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CG5964 PROTEIN.
GN GUG OR CG5964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
```

```
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003555; AF50413.1; -
DR Flybase; FBgn0020427; GUG.
DR InterPro; IPR001005; -
DR Pfam; PF00249; myb_DNA-binding; 1.
DR Pfam; PF01448; ELM2; 1.
DR SMART; SM00395; SANT; 1.
SQ SEQUENCE 1963 AA; 207927 MW; 8F386DE56E66CB68 CRC64;
```

Query Match 3.4%; Score 432; DB 5; Length 1963;

Best Local Similarity 19.0%; Pred. No. 6.4e-15;

Matches 437; Conservative 250; Mismatches 910; Indels 702; Gaps 100;

```
Qy 382 EIIDGLSEQENNEKQMRQLSVIPMMFD-----AQRRVKFINM--NGLMEDPMKV 431
Db 8 EIRVPGHQETDERELEESRSPGVADGDLMLFLRAARSMAAFQGMCDGLEDGLAAS 67
Qy 432 KDROFMNV-----WTDEKEIFKDKFTQHPKNFGL 461
Db 68 RDDTTINALDVLHDSGYDPGKALQALVKCPYSKIDDKWTEDEYKKFKLGRQKGFNR 127
Qy 462 I-ASVLERKSVDCVLYLYLTKK-----NEN-----YKALVRNRYGKRRNQIARSPQ 510
Db 128 IHKOLLPHKDPPELVFEFYLLWKTPGANNRRPHRRRSOSALRRNRVTRANNNSNTPPKK 187
Qy 511 EEKVE-----EKEEDKAETKEEKEEKEDEKESKENTREKDKID 555
Db 188 EDTPEPQTATTATAATAASRSPAVSKEENSSLTEDDASECSDSSLTTHKRDSP 247
Qy 556 G---TAETEEREQATPRGRKTANSQGRKGRITRSMTNEAAAA-----SAAAAAATE 605
Db 248 SRMTRNKQNNNSSTSGNNTAGNGGNATISGSGTGGAGAGNSSSKDOANAVANG 307
Qy 606 EPPPLPPPPPISTEPVETSRMTEEMEYAKGLVHEGRNMAAIKAVMGTKSEAKCNF 665
Db 308 KRPRGSETPDVSGASVDSPK-----TPTKAVAESSAN-----KRKGKQDTPNKK- 354
Qy 666 YFNKRRHNLNLLQHQKQTSRKPREERDVSQCESVAS----- 704
Db 355 ---KRTQESNEPSAHEENAIEKKRKPDP-SPEVSMNDSRDPVLDGDSNNTDTTTA 409
Qy 705 -----TVSAQEDDEDIEASNEEENPEDESEVAVKPSDENATSRGNTPEPAVELE 754
Db 410 EQQSTKDSKEIVSCKEEREMVTNLEAKAEKAIKAEALAEKSDSAIKNMDTEINTIQA 469
Qy 755 PTTTAPSTPSPLAVPSTKPAEDSVETQVNDISAEATQMDVDQOQSHAEESVCDPP 814
Db 470 SSATSLVDGPN---PNAIPSPAAPTVMKVTPTIATVETALNAAVDRKE-AIEKMECSD 525
Qy 815 PATKADSVDEVVRPENHASKEVDGNTKERDLDRASEKVE-----PRDEDLVAAQIINAQ 869
Db 526 PEMLKLATIKQEVSPQQOQHMOOSQOQQOQLAPGVIPQPPSPPPSESYII----- 578
```



















RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC024760; AAF59460.1; -  
DR InterPro: IPR003599; -  
DR SMART: SM00409; IG: 1.  
SQ SEQUENCE 6677 AA; 742580 MW; 7AC6B0B47BAll1987 CRC64;

Query Match 3.48; Score 429; DB 5; Length 6677;  
Best Local Similarity 18.88; Pred. No. 4.9e-14;  
Matches 500; Conservative 374; Mismatches 1016; Indels 772; Gaps 119;

QY 168 DQNASPKSLKEELIQSDMDRDRE-----IAKVEQIILKLLKKQOOLEEEAAKPPET 219  
DB 642 DESKKPSEVQPAEIVEQKDVVPQETSAPTVEKLAPVESKETPEVQAAEIVEQKDVVPET 701  
QY 220 EKPVSPPVQKHSRISVOIYDENRKKAEAKHIFE-----GLGPKVE--LPL 265  
DB 702 RAPTEVPTVE-KHTPV-----DSKETSEVPAEIVEQKDVVPETSAPTVEPTVEKHPTV 755  
QY 266 YN-----OPSDTKVYHENIKTNQVMKKLI-----LFFKRNHARKOREQKICQRYDOL 314  
DB 756 ESKSEVQPAEIVEQKDVVTCEEEIKELLTEVEVELEFF-----SQAEVFGLELDLL 807  
QY 315 MEAWKKVDRIENPRKAKESKTRREYKQFPEIRKQREQORFORVQGRGAGLSATIA 374  
DB 808 MECSEYVTTSIQKGSTAAPAQEP--VEKLAPVESKETSEVPAEIVEQKDVVPETSA 864  
QY 375 RSEHIEISIIQGLSEQENNEKQMQL-----SVIPPM--MFAEQRRVKFI 418  
DB 865 PSVEPTVEKLAPVESKETSEVQAAEIVEQKDVVPETSAPEVPTVEKLAPAESKETSEV 924  
QY 419 NNGLMEDPMKVKYKDRQPMNVTWDEKIFKDKFTQHPKNGFLGLASYLERKSVDPCLVY 478  
DB 925 QPAEIVEQ-KDVTCEEEIKELLTEVEVELEFF--FSQ-----AEVFGLELDLLMECSYV 975  
QY 479 YLTKNENYKALVRNRYGKRRG--RNOQIARPSOEKVEEKE-----EDKAEKTEKKEEKE 532  
DB 976 TTSIQKGSTAAPAQEPVEKLAPVESKETSEVQAAEIVEQKDVVPETSAPTVEPTVEKL 1035  
QY 533 KDEEKEKE--DSKENTEKD-KIDTAETEEREOATPRGKTANSQGRKRITRSM 590  
DB 1036 KPVESKETSEVQAAEIVEQKDVVPETSAPE-----PTVEKLAPVESKETSEVQ 1087  
QY 591 NEAAAAAATAAEEPPPLP-----PPEPI--STEPVETSRMTE-BEMEVAKKGLVYEH 643  
DB 1088 -----AEIIEQKDVVPETSAPTVEPTVEKLAPVESKETSEVQAAEIVEQKDVVP 1137  
QY 644 GRNMAAIKMWGTKEAOCKNFYFNYKRRHNLNLLQHQKQKTSRKRPR-----EERDVSO 699  
DB 1138 PETSAPTE-----PTVEKLAPVESKETSEVQAAEIVEQKDVVP 1177  
QY 700 ESVASTVSAQEDIEDASNEBENPEDSEVAVK--PSDSDPENATSRGNTPEAVE----- 752  
DB 1178 ETSAPTV-----EPTVEKLAPVESKETSEVQAAEIVEQKDVVPETSAPTVEPTVEKHAPV 1233  
QY 753 -----LEPTTET-APTGSTP-----SLAVPSTKPAE-----DSV 780  
DB 1234 ESKETSEVQPAEIVEQKDVVPETSAPTVEPTVEKLAPVESKETPEVQAAEILEQKDVTC 1293  
QY 781 ETQVNDIS-----AETAQOMQDVQ-----QEH---SAEEGSVCDP-----P 814  
DB 1294 EEEIKELLTEVEVELEFFSKAEVFGLELDLLMECSYVTTSIQKGSTAAPAQEPTEVKLA 1353  
QY 815 PATKADSDVDE-----VVRPENHASKVEGDNKTERDLDRASEKVEPDRDLDVAAQ 865  
DB 1354 PVESKETSEVPAEIVEQKDVVPETSAPTVEPTVEKLAPVESKETSEVQAAEILEQK 1412  
QY 866 INAQRPQSDNDSSATCSADEVDGEPERQRMFPMDSKPSLLNPTGSIILVSSPLKNPL 925  
DB 1413 V-----PVPET--SAPTV-----EPTVEKLAPVDSKE-----TSEVEPAE 1446  
QY 926 DLPOLQHRAAIVPMVSTPCNIPIGTPVSGYALYQRIKAMHESALLEQORQOEIDL 985

DB 1447 -----VEQKDVTC-----EIKELLTEVEV 1467  
QY 986 ECRSTSPCGTSKSPNREWEVLQAPHQLITNLPBGVRLPTTRPRPP--PPLISSKTT 1043  
DB 1468 ELLFSQAEVFS-----GLELDLLMECSYVTTSIQKG-----STAAPAQEPTEVKLA 1520  
QY 1044 VASKPSFIMGSGISOGTPGTILTSNQAQSYTOETPKPSVGSISLGLPQOESAKSATLP 1103  
DB 1521 TSEVEPAEIVE--QKDV-----VPETSAPTV-----EPIVE 1550  
QY 1104 YIKOEFSPRSQNSQOPELLVRAQHEGVYRGTAGAIQGSITRGTP-----TSKISVESI 1158  
DB 1551 KLKSVESKETSEVQAAE--IIEQKDVVPETSAPTVEPTVEKHAPVESKETSEVQAAE 1607  
QY 1159 PSLRGSITQGTIPALPOTGIPTEALYKGSISRM--PIEDSSPEKGREAAKSHVIEGSG 1217  
DB 1608 -----VEQKDVVPETSAPT--VEPTVEKLAPVE--SKETSEVEPAE--IVEQKDV 1652  
QY 1218 HILSYDNIKNAREGTRSPR--TAHEISIKRSYSEVEGNIKQGMRESPPV--SAP-LE 1270  
DB 1653 PV-----PETSAPTEPTVEKLAPVESKETSEVPAEIVEQKDVVPETSAPTVE 1702  
QY 1271 GLICRALPRGSHDLKERTVLGSGIMOGTFRATPESSEDGLKYPQIKKRESPPTRAPEG 1330  
DB 1703 PTIEKLAPVES-----KETSEVEPAEIVEQKDVSVPTSAPTVEP 1742  
QY 1331 AITGKPPYDGIITIKEMGRSIEIPRODILTOESRKTPEVQSTRIIEGSGISOGTPIKF 1390  
DB 1743 TIEKLAPVE-----SKETSEVEPAEIVEQKDVSVPE--TSAPTEPTVEKLAPVES 1791  
QY 1391 DNNS-----GOSAIKHNVKSLITGSKLSRGMPPLEIVPENIKVVER 1432  
DB 1792 KETSEVEPAEIVEQKDVVPETSAPTVEPTVEKLAPVESKETSEVQAAEIV----- 1842  
QY 1433 KYEDVKAAGET-----VRSRHTSVSSGSLRSTLHAPKAOLSGIYDD 1478  
DB 1843 -EHKDVQVPTSSPTVEPTVEKLAPVESKETSEVPAEIVEQK--DVPVPETSAPTVEP 1898  
QY 1479 TSARTPSYQNTMSRSGSPMNRSTSDVTIP-----PNKSNHERRKSTL 1521  
DB 1899 TVEKLAPVESKETSEVEPAEIVEQKDVVPETSAPTVEPTVEKLAPVESKETSEVQAAE 1958  
QY 1522 TPTQRESIPAKSPVPGVDVPSVSH--SPFDPHHRGSGTAGYVWSHLPTQLDPAFPHRA--L 1578  
DB 1959 VEQKDVSVPTSA--APTVEPTVEKLAPVE--SKETSEVQAAEIVEQKDVVPETSAPTV 2013  
QY 1579 DPAAYLQFQRLSTPGPSQYQIYAMENTROTILNDYITSQQMAY-----NLRPOVA 1632  
DB 2014 EPTV-----EKLAPVES-----KETSEVQAAEIVEQKDVVPETSAPTVEPTVE 2057  
QY 1633 RGLSPRE--QPLGLPYPATRGIIIDLTNMPPTILVPHGG--TSTPPMDRITYIPGTQITPPP 1690  
DB 2058 K-LAPVESKETSEVEPAE--IVEQKDV-----VPETSAPTVEPTVEKLA----- 2099  
QY 1691 RPYNSASMPGHTHAAAAAER 1749  
DB 2100 -PVESKETSEVQAAEIVEQKDVVPETSAPTVEPTVEKLAPVES-----KETSEVEPAEIV 2154  
QY 1750 SHGYVRSPPSVRTQETMLQORPSVFOGTNGTSVITPLD-----PTAQLRIMPLPAGGPSI 1805  
DB 2155 EQKDVSVPTSAPTVEPTVEKLAPV--ESKETSEVQAAEIVEQKDVVPETSAPTVEPTV 2212  
QY 1806 SOGLPASRYNTAADALAAALVDAASAPQDVSKTSKESHEARLEENLRSAAYSEQOQ 1865  
DB 2213 EKLAPVESKETSEVQAAEIVE-----QKDVTCSEEEIKELLTEVEVELEFFSQAEV 2266  
QY 1866 L-----EQKTEVEKRS-----VQCLYTSSAPFS-----GKQPHSS 1897  
DB 2267 LDLLMECSYVTTSIQKGSTAAPAHEPTVEKLAPVESKETSEVEPAEIVEQKDVVPET 2326  
QY 1898 VVYSBAGKDGKPPKSYRVEEELRTGKTITTAANFDVITITRIASDKDARERGSSSDS 1957



```
Db 2327 APTVEPTVEKLPVSKETSEVE-----PAEIVEQKDL----- 2359
QY 1958 SSSLSSHRYETPSDAIEVSPASS-----PAP-----POEKLQTOPEVVKANQ 2001
Db 2360 --PVETSAPTVEPTVEKLPVSKETSEVEPAEIVEQKDVVPETSAPTVEPTVEKLP 2417
QY 2002 AENDTQVEGLHHYRQOQESPPQQLPPSSQAEGMGQVPRHRLITLADHICQIITQ 2061
Db 2418 VESKETSEVEPA--EIVEQKDVVPETSAPTVE-----PTVEKLPVSKES----- 2459
QY 2062 DFARNQVSSQTPQ-----PTSTFQNSPVSALVTPVTKTSNRY----- 2102
Db 2460 ----KETSEVEPAEIVEQKDVVPETSAPTVEPTIEKLPVSKETSEVEPAEIVEQKDV 2515
QY 2103 --PESQAQSVH-----HORPGRSRVSPENLVDSKSGS-----RPGK 2135
Db 2516 SVPTSAPTVEPTIEKLPVSKETSEVEPAEIVEQKDVVPETSAPTVEPTVEKLPV 2575
QY 2136 SPERSHVSSEYPISSPPQV-----VHEKQDSLILLSQGAEPAE- 2177
Db 2576 SKETSEV--EPAEIVEQKDVVPETSAPTVEPTVEKLPVSKETS-----EVQPAEI 2626
QY 2178 -ORNDARSGSISYLPSTFKLENTSPMVSKKQEIFRKLNSGGGSDMAAQ----- 2230
Db 2627 VEHKDVQVPTSS--PTVEPTVEKLPV--VESK-----TSEVEPAEIVEQK 2670
QY 2231 ---PGTEIFNFI--PAYTTSGSVSSRSHGFADPASNLGLEDI--IRKALMGSEFDKVEDHGV 2285
Db 2671 VVPETSAPTVEPTVEKLPVSKETSEVEPAEIVEQKDVVPETSAPTVEPTVEKLPV 2730
QY 2286 MS-----OPMGV-----VPGTA-----NTSVTSGSTRREEDGDS 2315
Db 2731 ESKETSEVQPAEIVEHDKDVQVPTATTFTPTKEKLPVDSKETSEVQTAIVEQKDV 2790
QY 2316 PHSGGV----CKPLI----SKSNSKSKSPIGQGLTERPSS-----VSSVHSE 2359
Db 2791 PETSAPTVEPTVEKLPVSKETSEVQQAIVEQKDVVPETSAPTVEPTVEKLPVSK 2850
QY 2360 GDYHRTQPGWA--WEDRPSSTGSTOFFYPLNPLTMRLLSPPPTPIACAPSAVNOAAPHOQNR 2418
Db 2851 ETSEIQTAEIVEQKDVVPETSTSV-----VEPTKEKLPVSKETSEVQQA 2898
QY 2419 IWREPAELLSAQYETLSDSD 2440
Db 2899 IVEQKDVVPETSAPTVEPTKE 2920

RESULT 39
O43161 ID 043161 PRELIMINARY; PRT; 3851 AA.
AC 043161
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BASSON PROTEIN (K1AA0434) (FRAGMENT).
GN BSN OR K1AA0434.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Winter C., Dieck S.T., Bockmann J., Boeckers T., Kaempf U.,
RA Samarti-Vila L., Langnaese K., Altmann W., Stumm M., Soyke A.,
RA Wieacker P., Garner C.C., Gundelfinger E.D.;
RT "The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal
RT localization of the human BSN gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 2281-3851 FROM N.A.
RP TISSUE=BRAIN;
RC Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
```

```
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18448; CAA77176.1;
DR EMBL; Y18449; CAA77176.1; JOINED.
DR EMBL; Y18450; CAA77176.1; JOINED.
DR EMBL; Y18451; CAA77176.1; JOINED.
DR EMBL; AB007894; BAA23707.1;
FT NON_TER 1
SQ SEQUENCE 3851 AA; 410202 MW; 7CA91DB233284C161 CRC64;
```

```
Query Match 3.4%; Score 425.5; DB 4; Length 3851;
Best Local Similarity 18.8%; Pred. No. 3.6e-14;
Matches 568; Conservative 355; Mismatches 1132; Indels 967; Gaps 134;
```

```
QY 30 PNTRHQEFAYDPDRSSHLEVSQASQLLQOQQOQLRRPRLSLLSEFFHPSDRPQERRTSY 89
Db 458 PPTSQQPPVGAP--HRAS-----GTSPLKQKQPGQGLQPSGFL-----PAKASPLSTRKAS 506
QY 90 EPFHPGP--SPVDHDSLESKRPRLEQVSDSHFQVRAVLAFLVHPLPE----- 135
Db 507 LPSKASQAKPL--RASEPSKTPSSVQEKKT--RVPTKAEPMPKPPPTTTPATPKVS 562
QY 136 GLRASADAKDPAGGKHEAP-----SSPISQPCGDDONASPSKLSKEELIOSMDRV--- 188
Db 563 GVRRAEPA--TPVVKAVPEAPKGGAEADLVGKPYSDASRSPQSL--DTGYSSDGISS 618
QY 189 DREIAKVEQQLILKKKQOQLEEAAPPEPEKPVSPVPEQKHRSIVQIYIDENRKA 248
Db 619 QSETGVVQVQ-----EVEQLDSAGVTGPHPP---SPSEHKVGSSMRPLLAQOGLAPSE 669
QY 249 EAHKIFEGLGPKVPLYNOPSDTKVYHENIKTNQVMKKLILFFKRNHARKQREKIC 308
Db 670 RSKPLSSGTGEEQK---QRPHLSLITPEAFDSDEELEDDEDSAEWRREQQDTA 725
QY 309 Q-----RYDOLMAWEKKVDRIENPRKAKESKREYIEKQPEIKRQEQDERF 359
Db 726 ESSDDFGSQLRHDYVEDSSEGLSPLPPQPPARAALTDDEFMRRQILEMSAEEDNLEE- 784
QY 360 QRVCQRGAGLSATIA-----RSEHEISELIDGLSEQENNEKQMQLSVIPPMMDAQ-- 412
Db 785 DDTATSRGLAKHGTQGGPRPPEPSQEPALPKRRLPHNATTGYEELLPEGGSAGATD 844
QY 413 -----RVKFINMN-----GLMEDPMKYKDRQPMVNTDHEKIFIKDKF 452
Db 845 GSGTLQGLRRFKTIELNSTGSGYGHLDLQGGDPDS-----LDREPELEMESL 892
QY 453 IQHPKNFGLTASYLERKSVDCVLYYLTKNENKALVRNRYKRRGRNQOIARPSQEE 512
Db 893 TGSPED----RSRGEHSSTLPASTPSYTSSTPSLSEEDSDSPSRQRLEAKQOR 948
QY 513 KVEEKEDKAEKT--EKKEEKKDEEKKDEKSKENTKEKIDGTAETETEEREQATPR 570
Db 949 KARHRSGLPLPTIEDSSEELREELREELREELREELREELREELREELREELREEL 1001
QY 571 GRKTANSQGRKGRITRSMTNEAASAAAAAATEEPPLPPPEPEPEPEPEPEPEPEPEPE 630
Db 1002 DKEELRAQRRE-----RSKT-----PPSNLS--PIEDASPT 1032
QY 631 EEMEVAKGLVEHGRNWAATAKVGTKSEACKNFYENYKRRNLNLLQOHKQKTSRKP 690
Db 1033 ELQQAEE----- 1039
QY 691 REERDVQSCEASVASTVSAQEDIEDIASNEENPEDSEVAVKPSDESPENATSRGNTEPA 750
Db 1040 MEELHRSCESEYSPSPSL-----DSAEALDGGPSRLYKSGSEYNLPTF 1083
QY 751 VELEPTTET--APSTSPSLAVPTKPAED-----ESVETQVNDISIAE----- 791
Db 1084 MSLYSPETPTSGSSTTPSSGRP--LKSAAEAYEEMMRKAELLQROQQAAGARGPHGPSQ 1142
QY 792 -----TAEQMDVDQOQESHAEEGSCVCDPPPATKADSDVVEVRVPHENHASKVEDNTKE 843
Db 1143 PTGPRGLGSFEYQTTDREYQQAQPAAGTASLGAAYEEIL--QTSOSIVMRQASS 1200
```







RESULT 40  
ID Q9UPA5 PRELIMINARY; PRT; 3926 AA.  
AC Q9UPA5;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE NEURONAL DOUBLE ZINC FINGER PROTEIN.  
GN ZNF231  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99026125; PubMed=9806829;  
RA Hashida H., Goto J., Zhao N., Takahashi N., Hirai M., Kanazawa I.,  
RA Sakaki Y.;  
RT "Cloning and mapping of ZNF231, a novel brain-specific gene encoding  
RT neuronal double zinc finger protein whose expression is enhanced in a  
RT neurodegenerative disorder, multiple system atrophy (MSA).";  
RL Genomics 54:50-58(1998).  
DR EMBL; AF052224; AAC3555.1; -;  
SQ SEQUENCE 3926 AA; 416462 MW; 75CD56ADD437343 CRC64;  
  
Query Match 3.4%; Score 425.5; DB 4; Length 3926;  
Best Local Similarity 18.8%; Pred. No. 3.7e-14;  
Matches 568; Conservative 355; Mismatches 1132; Indels 967; Gaps 134;  
  
QY 30 PNTRHQEFVADYRSHLEVSQAOLQOQOQOQLRRRPSLLSEFHPGSDRPOERRTSY 89  
DB 533 PPTSQOPVGPAB-HRAS-----GTSPLKQKQKQGLQGPSGL-----PAKASPLTKASP 581  
  
QY 90 EFHFGCP--SPVDHDSLSKRRPLRQVSDSHFQRYSAALVPLVHLPE----- 135  
DB 582 LPSKASPAQKPL-RASEPSKTSPPSSQVQKKT---RVPTKAEPMKPPPTPTPATPKVKS 637  
  
QY 136 GLRASADAKDPAGFKHEAP-----SSPISGQPCDDQNASPSKLSKELIQSDRV--- 188  
DB 638 GYRAEPA--TPVVKAPPAKGAEDLVGKPYQDASRSQSL--DTGYSSDGISS 693  
  
QY 189 DREIAKVEQILKQKQOQLEEAAPPEKPYSPVPPVEQKHRSIVQIYIDENRKA 248  
DB 694 QSEITGVVQO-----EVEQLDSAGVTGPHPP---SPSEIHKVGSMPRLLOAQLAPSE 744  
  
QY 249 EAHKIFEGIPKVELPLYNQPSDTYVHENIKTNQVMRKLLIFKRRNHARKQEQKIC 308  
DB 745 RSKPLSSGTGEQK-----ORPHSLITPEAFDSDELEDLEEDSDSAEWRRRRQDDTA 800  
  
QY 309 Q-----RYDQLMEAWKVDRIENNRKAKESKTREYKQFPFIRKQEQQERF 359  
DB 801 ESSDDFGSLRDHYVEDSSEGLSLPLPQPPARAELTDEDFMRQILLEMSAEENLEE- 859  
  
QY 360 QRVGORGAGLSATIA-----RSEHIEISEIDLSQENNEKOMROLSVIPPMFDAQ-- 412  
DB 860 DDTATSGRLAKHGKQKQGPGRPRPEQSPAPALPKRRLPHNATTGYELLPEGGSAAETD 919  
  
QY 413 -----RRVKFTNMN-----GLMEDPMKYVKQRFNMVWTDHEKEIFKDKF 452  
DB 920 GSGTLOGGLRRKFTTELNTSGSYGHELDLQGGPDPS-----LDREPELEMESL 967  
  
QY 453 IOHPKNFGLIASYLERKSVPCVLYLYLTKKNENKALVRNRYKRRGRNQIARPSQEE 512  
DB 968 TGSPED----RSRGHSTLPASTPSYISGTSPTSLSLEEDSDSSPSRQRLLEAKQOR 1023  
  
QY 513 KVEEKEEDKAET--EKKEEBKDEEEDKEDSKENTYKDKIDGTAEEETEERQATPR 570  
DB 1024 KARHSHGPLPLPTIBDSSEEEELREELREQEKMRVEQOQRINRSTARKTR-----R 1076  
  
QY 571 GRKTSQGRKRGKTRITRSMNTNEAAAAAATAEPPPLPPPPPEPILSTEPVETSRWTE 630  
DB 1077 DKEELRAQRRE-----RSKT-----PPSNUS--PIEDASPT 1107

QY 631 EEMEYAKKCLVEHGRNWAIAKMWGKTKSEAQCCKNTFFNTYKRRHLDNLLQKHQKTSRKP 590  
DB 1108 ELRQAAE----- 1114  
  
QY 691 REERDVSOCEVASTVSAQDEDEIEASNEENPEDESEAVKPSDESDPENATSRGNTEPA 750  
DB 1115 MEELHRSCSEYSPSPL-----DSEALDGGPSRLYKSSSEYNMLPTF 1158  
  
QY 751 VELEPTTET--APSTSPSLAVSTKPAED-----ESVETQVNDISAE----- 791  
DB 1159 MSLYSPTTETPGSSSTTPSSGRP-LKSAEAYEEMMRKAELLQRCQQAAGARGHPGSPSQ 1217  
  
QY 792 -----TAEQMDYDQOEHSAEGSVCDPPPATKADSVDEVVRVPENHASKVEGDNKE 843  
DB 1218 PTGPRGLGSFEYQDTRREYQAAQPAEAEGTPASLGAAYEEIL--OTSQSTVMRQASS 1275  
  
QY 844 ROLDRASKEVPRDEDLVVAQOINAQRP--EPQSDNDSSATCSADEVDGEPEPERMFP 901  
DB 1276 RDLAFAEDKKKEK-----QFLNAESAYMDPMKONGGFLT-----PCTSTQLAAPV 1321  
  
QY 902 D-SKPSLLNPTGSILV-----SSPLKPN-----PLDLPLQLOHRAAV 936  
DB 1322 SFSTPTSSDSSGGRVDPVVRVTHFAKETQDPLKLHSSPASPSASKETGMFDSQPGT- 1380  
  
QY 937 IPPMVSCPCNIPICTPVSGYALYQRIKAMHESALLEEQRORQOEIDLECRSSTSPCGT 996  
DB 1381 -PATTAVAP--PAGLP-RGY-----MTPASPAGS 1406  
  
QY 997 SKSPN-----REWEVLQAPAHOLITNLPQEVRLP-----TTRP 1029  
DB 1407 ERSPSPSTAHSYGHSPITANYGSQTEQLPQAPSGLAAGRAAREKPLSASDGEQGTPOP 1466  
  
QY 1030 TR-----PPPLIPSSKTTVASEKPSFIMGSGTSQGTPTGYTLFHSNQAOSTQETPKPS 1082  
DB 1467 SRAYSFYASSPPLSPSS-----PSESPTF-----SPGKMGPRAT-----AEFSTQTPSPA 1512  
  
QY 1083 VGSISLGLPQOESAKSATPLVYKOEESPR-----SONSOPEGLLVRAQHEGVVR 1133  
DB 1513 PAS---DMRPSFGATPS--PWVAQGTQPHRPSTPRLVWQESSQEAFFMW-----ITL 1561  
  
QY 1134 GTAGAIQEGSITRGPTTSKI--SVESIPSLRGSITQGTPLPQTGIPTEALYKGSISRMP 1191  
DB 1562 ASDASSQTRMVASASTSPLCSPETQPTTHG-YQSTTPP-----SVSOLP 1606  
  
QY 1192 LEDSSPEKREAAKSHVIEGKSGHLSY-----DNKNAR-----EGTR 1233  
DB 1607 PEPPGP-PGFPRVPS-----AGADGPLALYGWGALPAENISLCSISSVPCTSRVEPGPR 1659  
  
QY 1234 SPRTAHEISLRSYSEVEGNI--KOGMSMRSPSVAPLEGLICRALP----- 1278  
DB 1660 TPGTA-VVDLRTAVKPTPIILTDQMDLTSIAVEARKYGLALDPIPGRQSTAVQPLVINL 1718  
  
QY 1279 RGSPHSDLKERTVLS-----GSMQGPTRATTFESFEDG-----LKYPKQIK 1319  
DB 1719 NAEHTFLATATTSITMASSVFNAQOPVYVGYQSRDLDFGOGGSGVCLAQKQVE 1778  
  
QY 1320 R--ESPPIRAEGAITKGPYDGIITIKEMGRSHEIIPQDIL----- 1360  
DB 1779 QAVOTAPYRSG-----PRGRPREAKFARYNLPNQVAPLARVDLITOMGTAOSIGLKPGPV 1834  
  
QY 1361 ----TQESRKTPPEVQVS-----TRP-----ITEGTSQGTPIKFDNNSG----- 1395  
DB 1835 PEPGAEPHRAFPALRSHALPGARKPHTVVVMQEGTACTVTTLLPEEPAGALDITGMRP 1894  
  
QY 1396 --QSAIKHNKSLITGPKSLSRGMPPLIIVPENIK----- 1428  
DB 1895 ESQACCDMWYKLPFG--SSCTGTETHPAPSVPEKSMADAAPPQSSSPFYGPRDPEPPPP 1953  
  
QY 1429 -----VVERCKYED-----VKAGETVSRHSTVSVSSGSPS----- 1457  
DB 1954 TYRAGGVVGPPEHQRYPQGLPGRLYSSMSDNTLAEAGLNYHAQRTGQLFOGPGGRDSA 2013



QY 1458 -VLRSTLHE-----APKAQLSPGI-YDDTSARRTP-----VSYQNTMSRSGSPMMNRTSDVTI 1507  
DB 2014 MDLSLKHSYSLGFGADGRYLGLOGLYSGVTDLRHPTDLIAHLPLPMRYSVSSVNIYSDHRY 2073  
QY 1508 PPNKSTNHERKSTL-----TPTQRESIPAKSPYGVDPVVSHPSPDPHHRGSTAGEVYWSH 1563  
DB 2074 GPRGDAVGFQENSLAQYSAITAREISRCAALNSMD-----QYGRHGSGGGGPDVLVOY 2127  
QY 1564 LPTQLDPAHPFRHALDPAAYLRFQRL-----SPTGCPYSQYQLYAMENR-----RQTILNDY- 1617  
DB 2128 QP-QHGPGLSAPQSLVPLRGLGNPTPEGHPSFGNLAQYGAQGTAVRQLLPSTAT 2186  
QY 1618 -----ITSQOMQVNLRPDYARG----- 1634  
DB 2187 VRAADGMIYTIINTPIAATLPTTTPASV-LRPMVRGMYRPYASGGITAVPLTSLTRVP 2245  
QY 1635 -LSPREQPLG-----LPYPATRGIIIDLNNPP-----TILVPHPGGTSTPPMD-- 1676  
DB 2246 MIAPR-VFLGPTGLYRYPAPSRFPJASSVPPAEGPVYLGKPAAKAGAGGSPRPEMPVG 2304  
QY 1677 -RITYIFGT-----QITFPPRPYNASMSPGHPTHLA-----AAASA 1712  
DB 2305 AAREEPLPTTPAAIKEAAGAPAPLAGQKP--PADAAPGGGSGALSRRPGFEKEEASQE 2362  
QY 1713 ERER-----ERER-----EKERERERIA-----AASDLYLRP----- 1740  
DB 2363 ERQRQQLQLQLERLEVELEKRLRLQLEELERELRERVELQRHREELIVORELQELQTI 2422  
QY 1741 -----GSEQPRGPGSHGVYRSPSPSVRQETMLQRPVSFQGTNGTSTVITPLDPTAQLRI 1795  
DB 2423 KHVVLQOOQEEQAQFALORQLAQRLQLEQIOLOQLOLQ-----QOLEEQQRQK 2474  
QY 1796 MPLPAGGPGSISQGLPASRYNTAADA-----LAALVDAASAPO----- 1833  
DB 2475 APFPAACEAPGPGPLAAELAQNGQYWPPLTHAFIAMAGPGLGQPREPVHLRGLPSS 2534  
QY 1834 -MDVSKTKESHEAARLEENLRSAAYSEQOLEQKT-----LEVERKRSVQC 1880  
DB 2535 ASDMSLOTQEEQWEASRSGIKKRHSRPLRDACELESGETPCVRRRIADSVQTDDEGES 2594  
QY 1881 LYTSSAFTSGRQPQPHSSVYSEAGKDKGPPPKSRVEELRTRGKTTIYAANFIDVITRQ 1940  
DB 2595 RYLLSRRRRARRSADCVSQTD-----EDSAEQEQVRRRRSR-----LPRH 2636  
QY 1941 IASDKDARERSQSDSSS-----SLSHRYETPSDAEIVSPA-----SSPAPP 1985  
DB 2637 SDGSDSKHDATASSSSAAATVRAMSSVGIQITISDCSVQTEPDQLPRVSPAIHITAADP 2696  
QY 1986 QEKLOTYQPEVVKANQAN-----DPTROYEG-----PLHHRP----- 2019  
DB 2697 KVEIVRYISAPEKTGRGESLACQTEPDQAQVAGPQLVGPVTAISPYLPGLQIVTPGPLG 2756  
QY 2020 -QOESPSF-----QOOLPPSSQAEGMG-QVPRHRLI-----TLADHICQIIITQDFARN 2066  
DB 2757 RFEKKKPPLEIGYQAHLPPELSQLVSRQPKSPQVLYSPVSPLSPH--RLDITSFSS 2814  
QY 2067 QVSSQTPQOP-----PTSTFQNSPVALVTPVTKTSNRYSPESQAQS 2109  
DB 2815 ERLNKAHVSPOKHFTADSALRQOITLPRPKTLQRS-----LSDPRLPSTAESAERESL 2870  
QY 2110 VHHQRP-GSRYS-----PENLVKSRGSRGKSPERSHVSSEPIEPISSQVQVVEHEKQDSL 2165  
DB 2871 YOHQGLGSOVSALPNPNSLVKRVKRTLSPPPPEAHL-----PLAQASQQLVAAS--- 2921  
QY 2166 LLLSQRG-----APPAQRNDARSPGISYLPSPFTKLENTSPMWKSKQELIFRKLNSG 2220  
DB 2922 --LQRLGTGTPVTPATKASLLRE-----LDROLRLVEHESTKLKRRQKQEL----- 2965  
QY 2221 GGDSDMAAQPCTEIFNLPAVTTSGSVS-SRHSFADPASNGLIEDIIRKALMGSDFDKV 2279  
DB 2966 --DEEKEIDAKLVEIGITQRKESLAKORGGDYPLRGLGEH---RDYLSSELNQL 3020  
QY 2280 EDHGVYMSQPMGV---VPGTANTSVVTSGETRREE---GDPSPH-----SGGVCK---PKL 2326

DB 3021 RLQG--CTTPAQGFVDFPATAAAPATPSGPTAFQOPRFOQPAPQISAGSGGPTQNGFFPAH 3078  
QY 2327 ISKNSRKSPI-----PGQGLYGTERRSSSVSHSEGDIHQTPCQWANEDEPSSGTGSTQ 2382  
DB 3079 QAPTYGPGSTYPAPAPPAGCASYPABEPLPNOQAFAFRPTGHIYAGOTP-----MPTTOSTL 3131  
QY 2383 FPNY-----PLTMRMLSSPTPTPIACAPSANQAAP-----HQ 2415  
DB 3132 FVPVADSRAPLQKPRQTSIADLEQKVPNTNYEVAS-PVPVMSASAPSETSYSGPVSSGYE 3190  
QY 2416 QNRWE-----REPAP 2426  
DB 3191 QGKVPEVPRAGDRGSVSQSPAP 3212  
RESULT 41  
Q9N3Y8  
ID Q9N3Y8 PRELIMINARY; PRT: 2344 AA.  
AC Q9N3Y8;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)  
DE Y4OC5A.3 PROTEIN.  
GN Y4OC5A.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID-6239;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RX MEDLINE-99069613; PubMed-9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RA Kalicki J., Randall J.;  
RT "The sequence of C. elegans cosmid Y4OC5A.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RC SEQUENCE FROM N.A.  
RP STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC024772; AAF60538.1;  
DR InterPro; IPR000194;  
DR InterPro; IPR002965;  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA; UNKNOWN\_1.  
SQ SEQUENCE 2344 AA; 7D8696D48D0ADDDB CRC64;  
Query Match 3.4%; Score 424; DB 5; Length 2344;  
Best Local Similarity 19.8%; Pred. No. 2.2e-14;  
Matches 457; Conservative 241; Mismatches 807; Indels 806; Gaps 109;  
QY 555 DGTAEETEE-----REQATPRGRKTANSQGRKRGRITRSMTNEAAAAAATAEPP- 607  
DB 173 DGDSTFAEIVLVTTEITTTTSETTTTSAITTEETTEPSTTE---PSTTEETTEET 229  
QY 608 -----PPPLP-----PPPEPISPEPVTSRWTE-- 630  
DB 230 TVSTNPPTPELSDOPTDGNATDIDDKIVKKEVSSYTAETPVAPPFFENTVTSVEEPPD 289  
QY 631 ---EEMEVAKKGLVEHGRNWAATAKWKYGTSEACKNFYFKRRHNLNLLQHKQ-- 684  
DB 290 DDAFEDSKPVKKSTLPNG-----INSKRST-----SYANPLTVDNAAESRWTK 333  
QY 685 -----KTSRKPREE--DVSOCSVASTVSAQEDIEDASNEENPEDESEVAVKP 733







ID Q9PU36 PRELIMINARY; PRT; 5120 AA.  
 AC Q9PU36;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE ACZONIN (FRAGMENT).  
 GN ACZ.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99439764; PubMed=10508862;  
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
 RA Klimmann M.W.;  
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
 RT zones, shares homology regions with rim and bassoon and binds  
 RT profilin.";  
 RL J. Cell Biol. 147:151-162(1999).  
 DR EMBL; Y19187; CAB60725.1; -;  
 DR HSP; P04410; I425.  
 DR InterPro; IPR000008; -;  
 DR InterPro; IPR001478; -;  
 DR InterPro; IPR001565; -;  
 DR InterPro; IPR002965; -;  
 DR Pfam; PF00168; C2; 2;  
 DR Pfam; PF00595; PDZ; 1;  
 DR PRINTS; PR00360; C2DOMAIN.  
 DR PRINTS; PR01217; PRICHTEXTENS.  
 DR PRINTS; PR00399; SYNAPTOTAGM.  
 DR PROSITE; PS00499; C2\_DOMAIN\_1; UNKNOWN\_1.  
 DR PROSITE; PS50004; C2\_DOMAIN\_2; 2;  
 DR SMART; SM00239; C2; 1;  
 FT NON\_TER 1  
 SQ SEQUENCE 5120 AA; 560760 MW; A658D9891B65B412 CRC64;

Query Match 3.3%; Score 415.5; DB 13; Length 5120;  
 Best Local Similarity 18.6%; Pred. No. 1.8e-13;  
 Matches 596; Conservative 380; Mismatches 1053; Indels 1183; Gaps 144;

Qy 3 SSGYPNQGAFSTEQSRYPHSHVQYTFNTRHQEFA-----PTKETPLASEKLGPMASDSTLTITKGSDEKK 803  
 Db 752 ASQAPPKEA---QAQPPKAA---PTKETPLASEKLGPMASDSTLTITKGSDEKK 803  
 Qy 40 ---VPDYRSSHLEVSQASQLLQOQQOQLR-RRPSLLSEPHGSDRDRPQERTSYE-----90  
 Db 804 PSLAKDSKHOTAEAKPAELSEQEAKSQPKVSCPLCKTGLNIGSKDPNPNFTCTECKKV 863  
 Qy 91 ---PFHGPSPVDHD-----SLEKPRLEQVSDSHFQRYSAAVPLVHPLPEGLRASA 141  
 Db 864 NLGCFNPMHIVEVQEWLNCOTQRAMSQGLGD---MGKV-----PLPK-LGPSQ 911  
 Qy 142 DAKKDPAPGGKHEAPSSPISQPCGDDQNASPSKLSKEELIQSDMDVDREIAKVEQ-----197  
 Db 912 PVSPPATPQKQPVAVSHSPQSSPTPTPAATK-PKEE-----PSVPKEVPKQQGKLE 965  
 Qy 198 QILKLLKKQOOLEEAAKPPPEPVPVSPPP-----VEQKHRSIVQIIVDENRKAEEAHK 252  
 Db 966 KTLSDAKTIQGIQKEDAKSKGKFLKFTPSADKIORVSKQEDSRLO-----1010  
 Qy 253 IFEGIGPKVPLYNOPSDTKVYH-----ENIKTNQVMKKLILFPKRNHARKQEQKIC 308  
 Db 1011 -----QTKLTKTPSSDKILHGKQEDIKFQEAKLAKI-----1042  
 Qy 309 QRYDQLEMAWEKKYDRINRPNRRKAKESKTRREYKQPEIRKQ---RQQRFORVFG---363  
 Db 1043 PSADKILHLRQKEDPKLQOMKMAKALSA-----DKIOPEAKQEDVQLQEVRLSKAVSD 1096  
 Qy 364 ---QRGAGLSATIAHSEHSEIISDGLSEQENNEKOMROLSVPPMFMFAEQRRVKFINM 421

Db 1097 KIOHGIQKDLNLQHVKIETKTSVEKIQEAKQESK-----1130  
 Qy 422 GLMED--PMKYKQKQFMNWTDEHEKEIFKQFIQHPKNFGLIASYLERKSVDPDCVLYYY 479  
 Db 1131 -LQDQKPLTISEDKIPATVSSDHKKLLSKSEEDKKPE-----LLEKSP-----H 1175  
 Qy 480 LTKKNENTKALVRNRYGKRRNRNQIARPSOEKVEE-KEED-----KAEKTEKKEE 531  
 Db 1176 PKDKEQITAEHTGHTGKVEVEAPCDKLHEKQEDVKEDLTGIGIQMWSKPEKAE 1235  
 Qy 532 K-----KDEEKDEKEDSKENTKEDK-----IDGTAETEER 564  
 Db 1236 KTPVPVSRPLPRSDHVEAVREKTEKEDDKSDTSSSQQKSPQGLSDTGYSSDGISSLSGEI 1295  
 Qy 565 EQATPRGRKTANSQGRKGRITRSM-----TNEAAASAAAAATEE-----606  
 Db 1296 PSHIPSDEKDLPREPSQKDTISQSPSPDLAKLESTVLSILEAQASTLTDEKSVKKE 1355  
 Qy 607 -----PPPLPPPPPEPISTEPVETSRWTEBEMEVAKKGLVEHGRNAAIAKM 653  
 Db 1356 LYETYEQTKDQHTKPLPVTPESSD-----BEDLEA-----IQEGERTTAADSK 1402  
 Qy 654 VGTKSEAOCKNFYFNK-----RRHNDLNLQHQKQTSRKPREERDVSOCEVA 703  
 Db 1403 GGASSQT-----DYKEEDGNDTPARRQRYDVEDSESSESPVPRKR-----A 1448  
 Qy 704 STVSAQED-----DIEASNESEN-----PEDSE-----VEAV 731  
 Db 1449 SVGSSSDYEYKRDSDGDEEDFIRKQIEMSADEDASGSEDDFIRNLQKEISVTSQ 1508  
 Qy 732 KPSEDSPE-----NATSRGNTPEPAVEPTTETAPST---SPSLAVPTKPAE 776  
 Db 1509 KKEEVKSAKGTGKRRMARKSAGYDAGRRSHWHDDODETFDESPEKYRTKSOD 1568  
 Qy 777 DESV-----ETQVNDNIS---AETAQMD---VQOQHSABEGSVCDPP-----814  
 Db 1569 GEELAISGGGLRRFKTIELNSTITSKYSETPEQOQGLYFDEEPELEMESLTDSPEDRS 1628  
 Qy 815 -----PATKADSVDEVVRVPENHASKVEGDNTER-----D 845  
 Db 1629 RGESSSLHASFPTPGTPTSVSSLDSDSPSH-KKLGESKQQRKARHSHGHLPLT 1687  
 Qy 846 LDRASEKVEPRDEDLVAAQINAQPEPQSDNDSSATCSADSDVDGDEPEPERQRMFMDSKP 905  
 Db 1688 IEDSSEEELEEEELKEQEKQREKQSSSKKKDKD-ELRAQRRERP-KTTP 1745  
 Qy 906 SLLNPTGSLVSSPLKPNPLDLPLQLOHRAAVIPPMVSTPCNIPIGTVPVSGVALYQRIK 965  
 Db 1746 SNLSPIEDASPTLELR-QAAEMEEL-HRS-----SCSEYSPSIESDPEGEISPEKII 1796  
 Qy 966 AMHE-----SALLEEQRQEQIDLECRSSTPCQTSKSPNREWEV 1006  
 Db 1797 EVQKYKLPATAVLSYPTDEKLGALKEESGK-----TLKSAEVEE 1840  
 Qy 1007 LQAPAPHQILNLPEGVRLPTRPTRPPPLPSPSSKTTVAESKPF-----IMGSGISQ---1059  
 Db 1841 MTHKTHK-----SKFQIASEKDEVEFEKESLYGGMLIEDY 1875  
 Qy 1060 -----GTPGTLYTSHNQAS--YTOE-----1077  
 Db 1876 IYESLIETNGTVDTNLARMQDESNEYIQORGKKEKIRASEQIYDEPKIFDLQEDYYS 1935  
 Qy 1078 -----TPKSVSGSISGLPQOE-----SAKSATLPYKOE--EF 1110  
 Db 1936 VEPLCSIVQEDIVSSVLIIPESHEIIVLDSTVTSTTEKQLLDAAAYEELMKQRQL 1995  
 Qy 1111 SPRSONSQPEGLLVRAGHEGVVRCAGIQEGSITRGTPTTKIS-VESIPSLRGSITQGT 1169  
 Db 1996 TPGSSPTQSDLAFTSDMKV--SSIGEIASTSLTSTTSIAISDVSSLSIALSI-----2049  
 Qy 1170 PALPOTGTPTEALVKGSIIRMPIDESSPEKGRREAAKSHVYEGKSGHILSYDNKKNAR 1229



Db 2050 ----PDVKI-----TQHTFAEIEDEYLTDAAREIQEIIISHETS--MLTYSEVS--- 2092  
Qy 1230 ECTSPRAHEISLKRSESVENIKQGMMSRESVSPAPLEGICRALPRGSPHDLAKR 1289  
Db 2093 EGAASILDATSLSSSTSSV-----CTDSSSPIDATGTV-----DTSDAVSK 2138  
Qy 1290 TVLGSIMOGTPRATSEFEDGLKYKPKQIKRSPPIRAFEAGITKPKYDGTITTIKEMGR 1349  
Db 2139 LVDSEDIIAQVFTSTEYSE-----VSMYVESVAGATTKPAIASDMDTVHOAAV 2188  
Qy 1350 SIHEIPRODILTOESKRKPEVVOSTRPIEGSISOGTPIKEDNNSGQSAIKHNKSLI-T 1408  
Db 2189 CLPE-----TAPSVFTTT--VIKPKQYASDTITVDISTAERKDAARKMKSTVET 2234  
Qy 1409 GSKL-----SRGMPLEIVPENIKVVERGKYEDVKAGET----- 1443  
Db 2235 GIUKTHEDSHKELSDMTRINLTGATSEQPLCVASVSVK-----EPASETPAVP 2285  
Qy 1444 ----VRSRHTSVVSSGP-----SVLRSTLHEAPKAQ-----LSPGI 1475  
Db 2286 TPRVVYKSTVSMPSAPALATSKVSLFRSSSLDSP-AQSPSPPPPPPPPPPPPLPPI 2344  
Qy 1476 YDDTSARPTVSYQNTMR-GSPMNRSDVTIPPNKTNHNR-----KSTLPTQRESIPA 1531  
Db 2345 LP-----KPAIYPRKKSQIOAPMATAPVPLVTSVATLESAAVLKNHVVPVTKTYTPT 2398  
Qy 1532 KSPVPGVDPVYSHSPDHRGRTAGEVYWSHLPTQL-DPAMPFHRALDPAAAYLFFORQ 1590  
Db 2399 PPPVP---PKFSSIP-----AGLVF-SHRPTEVTKPIAPKPAVPPPLPIAVHKPAE 2445  
Qy 1591 LSPTPGYPQYOLYAMENTROTILNDYITSOQMIV-----NLRPDVARGLSPREOPLGLP 1645  
Db 2446 TQPKP-----IGLSLTSMTLN-LVSSAEYKTASTPSLSPHSNKS-SPR---LTKP 2492  
Qy 1646 YPATRGIIDLNMPP--PILVPHPGGTGP-----PM----- 1675  
Db 2493 SOETVWITLSEPGTPTTEAITSOAVTSWPLEAPSKEIQIOPMQPIFTSSMKAVEIQSMA 2552  
Qy 1676 DRITYIPG-----TOITPPRPYNASMSGPHTHLAAAASAEERE----- 1717  
Db 2553 DOSMITGALQITPITTOSTFEKVP---SSKSEAVTTEVAKTTASVVRKVPVPSVGLSVT 2609  
Qy 1718 ----RERERERER-----IAAASSDLYL-R-P 1740  
Db 2610 ITIPPEIYISQPYRENGRPHGLGDVLDLTLTKVDIEMRSCMDLSAVSMDARROMP 2669  
Qy 1741 GSEQGRG-----SHGYVRSPSPSVTQTMLO---QRPSVQGN-----GTS 1782  
Db 2670 TSDTSGRPVSTVQPAIINLSTACVADPSLSIVTETVATVCTATVSYGASTDSLVDLGH 2729  
Qy 1783 VITPLDPTAQLRIMPLPAGGPSISQGLPASR-----YNATAADALAVDAAASAP----- 1832  
Db 2730 MTTPLQLTTSKHF--PAYRTVSQPFVSRDEVPINLSGTSAAVTAATKPVTPVPV 2787  
Qy 1833 ----QMDVSKTKESKHEAARLENLRSLRASAASAEQOQLEQKTLV 1873  
Db 2788 SVTNGWTDLSSTQEPMEIGAVDLSTTK-SHRTVVTMDTETSGIITTVIEDDE-KPVDLTA 2845  
Qy 1874 EKRVSQC-LY-----TSAPFSGKQPH-----SSVYVSEAG-----KDKGPPPKSR 1914  
Db 2846 GRRVACCDVMVYKLPFRSCTAQOQPTTLPEDRFGYDRSDGSGYGRGMGMKPSM 2905  
Qy 1915 YEEELRTGKTTIAANIDVITROIASDKDAREGSSQSSSLSSSHRYETPDAIE 1974  
Db 2906 SDTNLSEAGLFAYKSKNSFDYQV-----GATDAADVLTSGRV-TSGEVM 2949  
Qy 1975 VISPASSAPPOEKLOTVQPEVVKANQANENDPTQYEGPLHRYRQESPSPOQLPSS 2034  
Db 2950 YSSKTTGYP-----ETRQVISGIGISTPQY---SQARWVSSLS 2985  
Qy 2035 QAEGMGQVPRHRLITLADHICQIITQDFARNQVSSQTPQOQPPSTFQNSFALVSTPVR 2094  
Db 2986 SPFGAGSVLRSSNGV-----YSSVATPIPTFAITQPGSIFSTVR 3028

Qy 2095 ----TKTSNRYSPESQA--OSVHHORPGS-----RVS 2120  
Db 3029 DLPTLOTIDSVPSLSTQONQPLPRYSFLTTMAEKDASTTLDIETGLPPLTLESIAE 3088  
Qy 2121 PENLVKSRGRSPKSPERSHVSSEPEYPIPPQVPVY---HEKQDSL---L 2166  
Db 3089 PTNLIPAT-----ASEVYTDVIEDEVALIIAPECKQOQLDLERELLELEK 3134  
Qy 2167 LLSQRGAPAE-QRNDARSQGISVLPSSFYTKLENTSPMWKSKQEI----- 2212  
Db 3135 IKOORFAELEWEQEIQR-----FREOEKF--MVOKLEELQSMKHLLFQOEE 3182  
Qy 2213 ----FRKL-----NSSGGG----- 2222  
Db 3183 ERQAYMMRQETLAAQOQLQLEQFOOQLOOLHOOLEEOKIROIYQYGYDPSGTGSPQTMTD 3242  
Qy 2223 ----DSMAAQAQGTIEFNLPVATTSGS-----VSSRG-----HSFAD 2256  
Db 3243 QALLEGQYATAENG-QFWPTDDATTASGVLGIEISQSQTWVTVQSDGITOYIPRSGILS 3301  
Qy 2257 PASNLGLEDI-----IRK-----ALMGSEDD-----KVEDHGVMVMSQPMG 2291  
Db 3302 SVSEMSLKDIDVREEKQLKRSSMPKLRGYEELEESLEEEPRCYKKIVDSGVQTDDEG 3361  
Qy 2292 VVPGTAN-----TSVVTSGETREEGDPSPHSGVCKPKL--ISKNSRKS- 2337  
Db 3362 ADRGVTNRRRTKKSVDTSVQTDDED--ODEWDLSSRGR--KPRVGKSESTTEADKAKQ 3418  
Qy 2338 ----PI-----PGQYLG----- 2346  
Db 3419 FSKVSSIAQVVAEISVQTEPVGIRTPSIRARLDAKVEIIKHISAPEKTYKGESLGCQT 3478  
Qy 2347 ----TERPSVSVSHSEGDYHROTP---GWAWEDRPSSTGSTOFFYPPNPLTMRMLSSPPP 2398  
Db 3479 ETESDTOSQOYLSASSPOKDKRPTPLEIGYSSHLRDPDSTLQV-VPSPPKPKVLYS-PI 3536  
Qy 2399 TPIACAPSANVQAA--PHOONRIWEREPAPLL 2428  
Db 3537 SPV--SPSKVIESAFVPEYKSTIDDISPOKML 3566

RESULT 43  
Q9H4R4  
ID Q9H4R4 PRELIMINARY; PRT; 102 AA.  
AC Q9H4R4; 2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE BB329D4.2 (NOVEL PROTEIN SIMILAR TO A TRUNCATED NUCLEAR RECEPTOR CO-  
REPRESSOR 1 (NCOR1) (RETINOID X RECEPTOR INTERACTING PROTEIN 13)).  
GN BB329D4.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Babbage A.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL391119; CAC14275.1; -  
KW Receptor.  
SQ SEQUENCE 102 AA; 11336 MW; 462EC7CAAE382014 CRC64;

Query Match 3.3%; Score 414.5; DB 4; Length 102;  
Best Local Similarity 43.7%; Pred. No. 1e-15;  
Matches 90; Conservative 3; Mismatches 4; Indels 109; Gaps 1;  
Qy 1 MSSSGYPNQGAFSTESQRYPPHPSVOYTFPNTHQOEFAVPDYSRHLEYSQASQLLOQQ 60  
Db 1 MSSSGYPNQGAFSTESQRYPPHPSVOYTFPNTHQOEFAVPDYSRHLEYSQASQLLOQQ 36



```

Qy 61 OQQQLRRRPSLLSEPHGSDRPOERTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQ 120
Db 37 ----- 36

Qy 121 RVSAAVPLVHPLPGLRASADAKDPAGFKHEAPSPISQPCGDDQNASPSKLSKEE 180
Db 37 -----DPAGFKHEAPSPILGQPCGDDQNASPSKLSKEE 71

Qy 181 LIQSMRDVREIAKVEQOILKLLKKQ 206
Db 72 LIECDRDVREIAKVEQOILKLLKKQ 97

RESULT 44
P70039
ID P70039 PRELIMINARY; PRT; 2829 AA.
AC P70039; P79934;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ADENOMATOUS POLYPOSIS COLI.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Vlemickx K., Wong E., Guger K., Gumbiner B.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U64442; AAB41671.1; -.
DR HSSP; Q02248; 3BCT.
DR InterPro; IPR000225; -.
DR Pfam; PF00514; Armadillo_seg; 6.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 2829 AA; 310879 MW; 8A2BABDB7706E496 CRC64;

Query Match 3.3%; Score 412; DB 13; Length 2829;
Best Local Similarity 18.6%; Pred. No. 1.3e-13;
Matches 480; Conservative 330; Mismatches 880; Indels 890; Gaps 116;

Qy 169 QNASPSKLSKEELIQSMRDVREIAKVEQOILKLLKKQOOLEEA----- 213
Db 759 OKALEAELDAQHLSETFONIDNLPKTHH-NKORHKQNLCEYALDSSRHDDSIICRSN 817

Qy 214 -----AKPEPEKVPSPPPVEQKHSRIQIYIDENKKAEEAHI 253
Db 818 FSIQNLTVLSPYINTVLPGSSPRPTWDGSRPEKDRERTAGLGNVHTTSSGNSKRI 877

Qy 254 FEGLPKVEL-PLYNQPSDTKVYHENIKTNQVMRKLLILFFKRRNHARKQEQKICQRYD 312
Db 878 GIQLSTTAQISKVMDVSNHILVQENSSGSA-SEMHCMDSERNRQKPPS----- 927

Qy 313 QLMEAWKKVDRIENNRKAK-ESKTR-----EYKEQ- 345
Db 928 -----NHQPSNPFTTKAESSTRGCPVAFMKMEYKMAINDSLNSVSTEGYKRG 977

Qy 346 --FPEIRKQEQE-RFORVQORGAGLSATARSEH-----EISII 384
Db 978 QVKPSVESYSEDDSKFFSYGQYAGLAHKIOSANHHMDNDTELDTPINYSLKYSDEQLN 1037

Qy 385 DGLSEQENNEKQRLSVIPPMFADAEQRRVKFINMGLMEDPMKVYKDRQFMNVWTDHE 444
Db 1038 SGROSPQNEWRSPKHIIDSEMKQEQORP-----TTKTYTSYITENK 1082

Qy 445 KEIFKDFIQHPKNFLGFIASYLERKSVPCVLYLYTLTKKENYKALVRRNYGKRRNQ 504
Db 1083 EEKHK-KF--PPHF-----NQSENVA-----YTRSGANNO 1111

Qy 505 I-----ARPSQEK-----VEKEEDKAEKTEKKEEKDEKDEKDEKENTKEK 551
Db 1112 VDQSRVSNLNSKASKPHCOVDYDDDKTTF-----SERYSEEQEQEDETQRN---K 1164

```

```

Qy 552 DKIDCTAETEREQATPRGRK-----TANSOGRKRGTRITRSMTEAAAA 596
Db 1165 YNIKAYASEEHGGEOPIDYSRKYSTDVPSSAQKPSFPYNNSSKOKPKKEQVSSNSNTPT 1224

Qy 597 SAAAA-----AATEP-----PPPLPPPPPEISTEPVTSRWTEEMEYA-KGLGYE 642
Db 1225 PSPNSNRQNLHPNSAQSRPGLNRPKQIPNKPSPINQETIQT--YCVEDTPICFSRG--- 1279

Qy 643 HGRWAAATAKWGTGKSEAQCKNFYNYRRHNLNLLQOHKQKTSRKPRERDVSQCESV 702
Db 1280 -----SSLSLSSAEDEIEGRE--RNSRGQESNNTLQ-----ITEPKISAVSKDGAV 1325

Qy 703 ASTVSA-----OEDIEDIASNEENPEDEVAVKPSESPENAT--SRGNTEPAVELEP 755
Db 1326 NETRSSVHHTRTKNRLQTSN-----ISPDSSRHKSVEFFSGAKSPKSGAQ 1373

Qy 756 TTETAP-----STSPSLAVPSTKPAEDEVET-QVNDISISAEATAEQMDVDQOBS----AE 806
Db 1374 TPKSPPEHYVQETPLMFSCRCTSGSLDSFESHSIASSVA-----SEHMISGIIS 1426

Qy 807 EGSVCDPPPATKADSDVDVEVRVPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOI 866
Db 1427 PSDLPDSPGQTPMPPSRSKTPPPQTVQAKDG-----SKPIVPDEERGKVAK-- 1473

Qy 867 NAQRPEQSDNDSSATSADSDVDCGEPEQRQ--FPMDSKPSLLNPTGSIILVSSPLKPNP 924
Db 1474 -----TAVHSAIQRVQLQADTLHLHATSTPD-----GFSCASSLSALS 1514

Qy 925 LDLPOLQH--RAAVIPPMVSTPCNIPITGTPVSGYALYORHIKAMHESALL-----EEQ 976
Db 1515 LDEPYIQDVQVLKIMPPVLENDQGN-----KAEPEKEFIDNKAKEDEK 1557

Qy 977 RQREQO-----IDL--ECRSSTSCGTSKSPNRWEVLQAPAPHLITNLPCEVRLP 1025
Db 1558 RSEQEKMDLDDTDDIDILEECIISAMP-----RKPSR-----KNKKVP 1596

Qy 1026 TTRPTRPPPL-----IPSSKTTVASEK-PSFTMGSGIS-----QGTPTGYTYS 1068
Db 1597 QPTPGKPPPPVARKPSOLPVYKLLSSQNRLQTKHVNTSHDDMPRVYCVETGPINFSTA 1656

Qy 1069 HNAQSYTOETP-----KPSVGSISLGLPROQESAKSATLPYIKOEFPSPKSNQSQBEG 1121
Db 1657 TSLSDLTIESPPSEPTNDQNTDLSLTDLERD-----TIP----- 1692

Qy 1122 LUVRAQHEGVVGTAGAIQEGSITRGTPTSKISVESIPSLRGSITQGTAPALPQTGIPTEA 1181
Db 1693 -----TEGRSTDDTDASK-----PLNPTTVLDEDK 1717

Qy 1182 LVKGSISRMPIEDSDPEKREAAKSGHVIYBKGSGHILSYDNIKNAREGTRSPRTAHEI 1241
Db 1718 AEEGDILAECIHSAMPK-----GKSHKPYRVKK--IMDQINHITSAAATSSGNSRSMQET 1768

Qy 1242 SLKRSYEV-----EGNIKQGMRESVPVAPLEGICRALPRGSPHS----- 1284
Db 1769 DNKKTSPVKPMQPSIGFKEKRLKNTLKNLPS--ENQYCD--PR-KPSSKKPKSVAN 1822

Qy 1285 ---DLKERTVLSG-----SIMQCTPRATTES-----FEDGLKYPKQIKRESPP 1324
Db 1823 EKIPNNEERT--KGFAFDSPHHYTPTEGTPYCFSRNDSLSLDFEDD---DIDLSEKAE 1877

Qy 1325 IRAFEGAITKGPYDGIITIKEMGRSIIHEIPRODILTQES---RKTPE-VVO----- 1372
Db 1878 LRKEKGT---KDTQKVKYKHENRAINPMGKQDQTPKPSLGGROPKALVQKPTSFSSA 1933

Qy 1373 -----STRPIIEGSIQGTPIKFDNNSQSAI-----KHNYKSLITGFSKLSRGM 1418
Db 1934 AKGTQDRGADKDEKMFENAIENTPVCFSRNSSLSDIDQENNKE--TEPLKOT---- 1987

Qy 1419 PLEIYPENIKVVERGKYEDVAGETVRSRH---TSVSSGSPVLSSTLHEAPKAOLSPGI 1475
Db 1988 -----GTSETOLGRRPQTSYGAPKPSFHVEDTPVCFSRNSSLSDIDSDDDLLQECI 2040

```



```
QY 1476 YDTSARPTVSYQNTMGRG-----SPMNRTSDVTIPPNKSTNHERKSTLTPTQRESIP 1530
Db 2041 SSAMPKKRPSKIKNEVCKSRNSVGGILAEEDPDLTLDRIOQSDSENAFSP-DSENF 2099
QY 1531 AKSPVGVDPVSHSPFDPHHRGSRAGEVYWS-----HLPTQLDPAMPFHRALD-- 1579
Db 2100 WKAIQEGANSIVSRLE-----HQAAGSLSROGSDSDSILSKGSGISLGPFLTLDK 2154
QY 1580 -----PAAAYLFORQLSTP-----GYPSQYQLYAMENTROTILNDYIT-- 1619
Db 2155 EKTITSNGPKILKPAEKSALENAKTEEPKIGKGVKYSKITGKRSKSS--SDFSSHC 2212
QY 1620 SOOMVNLRPDVARG-----LSPREOPLGLPYPATRGIIIDLTNMPPTILVPHPG 1668
Db 2213 KOSVOTNM-PSISRGRTMHIHIGVRASSPSTSPVSKKGPVK-----NVPSKSGNEPS 2265
QY 1669 GSTTTPMDRITVPGTQITFPPRPYNASMSGPHTHLAAASABREREREKERERER 1728
Db 2266 SSSSP-----KGT-----KPLKSELVYGRSPSTPGGSKGNSRSGSRD----- 2304
QY 1729 IAAASDLYLRGSEOPGRPGSHGVVRSPSVRTQETMLQORPSVFOGTNGTSTVITPLD 1788
Db 2305 --SASS-----RSPSOPLSRP-----LQSPG-----RNSISPKNG--ISPPN 2338
QY 1789 PTAQLRIMPLPAGGPSISQGLPASRYNTAADAALAALVDAASAPQMDVSKTESKHEAAR 1848
Db 2339 KFSQLPRTTSPSTASKSG----- 2358
QY 1849 LEENLRSAVAEVOOQLLEKQLEKRVKSVQCLYTSSAPPGSKQPHSHSVVYSEAGDKG 1908
Db 2359 -----SGRMSYTSRQOLSQPNLSKQ-----SGLPKTHSSIPRESAS-- 2396
QY 1909 PPKRYEEELTKTKTITTAANFIDVITROIASDKDARERGSSDSSSLSSHRVET 1968
Db 2397 -----KSLNQNTGNS-----KKV-----ELSRMSSYKSGSESDRSE 2431
QY 1969 PSDAIEVISPASSAPPOEKLQYQYEVVYKANKAENDPTROYEGPLHYRPOQESPSP-- 2026
Db 2432 PA-----LVROSTFK----- 2449
QY 2027 QOOLPPSSQAEQMGVPRTHRLITLADHLCIITODFARNQVSSQTPOOPPTSTFQNSPS 2086
Db 2450 RKLEESASFELS-----SSSRADSPRSTQTPA 2480
QY 2087 ALVSTPVRTKTSNRYSPESQAQSVHH-ORPGSRVSPENLVKSRGSRPSPERSHVSSE 2145
Db 2481 LSPSLP-----DMALSTHSIQAGGWKPPNL--NPAEHGDSRRHDISRS 2525
QY 2146 PYEPTSPQVPPV-----HEKQDSLLLSQRGAEPAEQNDARSFGSISYLPSTFTK 2197
Db 2526 HSE--SPRLPITRSGTWKREHSHKSSSL-----PRVSTWRTTSGSSSILS-- 2569
QY 2198 LENTSPMKVSKQEIFRKLNSGGGSDMAAQAQCT-----EFNLPVAVTSGSVSR 2250
Db 2570 -ASSESSKASEDEKQVCSPGPRSCSSAKGTWKRIESELTP-----SN 2619
QY 2251 GHSFADPASNGLIEDIRKALMGSDDDKVEDHGV-----VMSQPMGVVGTANTSWVTS 2305
Db 2620 GSSSTIAESNCLESKTLVYQMAPAVSKTEDVWVRIEDCPINNPSRSGSPGNSPPVIDN 2679
QY 2306 -----ETRRREGDSPH-----SGVCVKPLKLSKNSRKSQSPGQGYLGTERTSPSSVSHS 2358
Db 2680 VLDQCKEAAKDCHTRNSNGNVPFL--ENROKSIKVDGLDTKTGD-PKSLINQQ 2735
QY 2359 EGDYHROTPGWAWDRPSTGTS-----TOPPYNPLTRM-----LSSTP--PPI 2401
Db 2736 ETNENTVAERTAFSSSSSKSHSPSGTVAARVTPNYPNPRKSGNENSTSRPSQIPTPV 2795

RESULT 45
Q94226 PRELIMINARY; PRT; 2361 AA.
ID Q94226
AC Q94226;
```

```
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE SIMILARITY TO SALMO GAIADNER.
GN F45E4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilson R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U70852; AAB09135.1; -.
SQ SEQUENCE 2361 AA; 261579 MW; 75F7106658D38E3B CRC64;
```

Query Match 3.2%; Score 409.5; DB 5; Length 2361;

Best Local Similarity 18.6%; Pred. No. 1.4e-13;

Matches 523; Conservative 374; Mismatches 975; Indels 943; Gaps 138;

```
QY 79 SDRPOERTSYEPHPGPSVDHDSLESKRPLEQVSD--SHFORVSAAVPLVHPVPEG 136
Db 22 SDRPSALAMF-----TNPSASPSL--LRKESDRGDIILNLRSSGA-----D 65
QY 137 LRASADAK---KDPAGFGKHEAFSPSPISGQPCGDDQNASP-----SKLSKEBILQSM 185
Db 66 SRASNDSSASRLPDTALVGLSEAEKQHTMSVMSRNRNTSPMTSRRCSSALQMLPEVDNL 125
QY 186 DRVORE-----IAKVEQO-----ILKLKKQO-----QLEEEAAKPPPEKP 222
Db 126 SEAEKEHTQTLKAEKSTPFPMKVPKMKQISSRTESNRSVSSGDEEVENEVQRKT 185
QY 223 VSPPPVEQKHRSIVQIYDENRKKAEAAHIFEGLPKVEL----PLYNQPSDQTVHYHEN 278
Db 186 IEEPIVEIPSAVT---PRNLRV-----IPPIAISHPTTPHSAKTDGSRHS 232
QY 279 -----IKTNQVWRKKLILFFKRNHARKQROKICQRYDQLMEAWEKVDRINNP 329
Db 233 GSSAHSQFGFTSPISGFK--IFFDK--AKTATETLKEIKDEIVIVDK--DKTETP 285
QY 330 RKAKESTREYKEQFPIRKOQERQORFQVQORGAGLSATI----ARSEHEITSEID 385
Db 286 EPNVSNELTAE-----EFEHIRINEM-----AGIDEPIOPPIAERRKSSVVS 330
QY 386 GL-----SEQENNEKO-MRQLSVIPPMFDAQRRVKFINMGLMDPMKVYKD 433
Db 331 GLKNMFGVGKHDESELTTEKEHIRMKS-----LMAEKM----- 364
QY 434 RQFMVNTDHEKEIFKDKFIQHPKN-FGLIASYLERKSVPCVLYLYLTKKNTYKALVR 492
```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:39:33 ; Search time 19.61 Seconds  
(without alignments)  
4262.284 Million cell updates/sec

Title: US-09-522-753-11

Perfect score: 12643

Sequence: 1 MSSSGYPNQGAFSTEQSRY.....EREAPLISAQVETLSDSD 2440

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	12643	100.0	2440	1	NCRI_HUMAN
2	11536.5	91.2	2453	1	NCRI_MOUSE
3	4147.5	32.8	2517	1	NCRI_HUMAN
4	4055.5	32.1	2472	1	NCRI_MOUSE
5	2493.5	19.7	533	1	NCRI_RAT
6	480	3.8	2468	1	MAPB_HUMAN
7	455.5	3.7	2774	1	MAPA_RAT
8	447	3.5	2464	1	MAPB_MOUSE
9	437.5	3.5	2805	1	MAPA_HUMAN
10	437	3.5	2845	1	APC_MOUSE
11	411.5	3.3	3924	1	ANK2_HUMAN
12	406	3.2	2842	1	APC_RAT
13	404	3.2	3256	1	K167_HUMAN
14	390	3.1	1226	1	YCS3_YEAST
15	374.5	3.0	3562	1	PGCV_CHICK
16	365	2.9	2843	1	APC_HUMAN
17	360	2.8	2349	1	TPR_HUMAN
18	353	2.8	2142	1	BAT2_HUMAN
19	350	2.8	661	1	YDBJ_SCHPO
20	348.5	2.8	1018	1	YKZ6_CAEEL
21	343.5	2.7	1411	1	TCOF_HUMAN
22	343	2.7	3866	1	HRX_MOUSE
23	336	2.7	2688	1	ZEPL_MOUSE
24	332.5	2.6	1861	1	MAP2_RAT
25	326	2.6	1781	1	AKAC_HUMAN
26	321.5	2.5	3969	1	HRX_HUMAN
27	317	2.5	4687	1	PLEI_RAT
28	314.5	2.5	3358	1	PGCV_MOUSE
29	312	2.5	1902	1	SMEL_HUMAN
30	310.5	2.5	2004	1	MOZ_HUMAN
31	309.5	2.4	2441	1	CBP_MOUSE
32	308	2.4	2442	1	CBP_HUMAN
33	303	2.4	3396	1	PGCV_HUMAN

34 300.5 2.4 2090 1 HFCL\_MESAU  
35 299.5 2.4 2715 1 TRX2\_HUMAN  
36 298.5 2.4 771 1 CALD\_CHICK  
37 298 2.4 2194 1 SC16\_YEAST  
38 296.5 2.3 1828 1 MAP2\_MOUSE  
39 295.5 2.3 848 1 NFM\_MOUSE  
40 295.5 2.3 2035 1 HFCL\_HUMAN  
41 295 2.3 2717 1 ZEPL\_HUMAN  
42 294 2.3 915 1 NFM\_HUMAN  
43 291 2.3 1827 1 MAP2\_HUMAN  
44 289 2.3 1637 1 MRSP\_STAAU  
45 284.5 2.3 1505 1 CDP\_HUMAN

## ALIGNMENTS

RESULT 1  
NCRI\_HUMAN  
ID NCRI\_HUMAN STANDARD; PRT; 2440 AA.  
AC 075376; Q9UPV5; Q9UQ18;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).  
GN NCOR1 OR KIAA1047.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;  
RA "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses  
RT transcription by interaction with the human N-Cor/msin3/HDAC1  
RT complex";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).  
RN [2]  
RP SEQUENCE OF 782-2440 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=9397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
Tanaka A., Kotani H., Nomura N., Ohara O.;  
RA "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
RN [3]  
RP SEQUENCE OF 974-2440 FROM N.A.  
RX MEDLINE=9375332; PubMed=1044336;  
RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,  
Horwitz K.B., Lupski J.R., Seo H.;  
RA "Localization of the human nuclear receptor co-repressor (N-COR) gene  
RT between the CMT1A and the SMS critical regions of chromosome  
RT 17p11.2";  
RL Genomics 59:339-341(1999).  
RN [4]  
RP FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
CC ABSENCE OF LIGAND.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
CC DOMAINS (ID1 AND ID2).  
CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND



CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
CC SPECIFICITY.  
CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A-SANT-B).  
CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.  
CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF044209; AAC33550.1; -;  
CC EMBL; AB028970; BAA82999.1; -;  
CC EMBL; AB019524; BAA75814.1; -;  
CC MIN; 600849; -;  
CC InterPro: IPR001005; myb.DNA-binding; 2;  
CC Pfam: PF00249; myb.DNA-binding; 2;  
CC PROSITE; PS50090; MYB\_3; 1;  
CC Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
KW Coiled coil.  
FT DOMAIN 174 216 COILED COIL (POTENTIAL).  
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B.  
FT DOMAIN 299 328 COILED COIL (POTENTIAL).  
FT DNA\_BIND 437 482 SANT-A (POTENTIAL).  
FT DNA\_BIND 625 670 SANT-B (POTENTIAL).  
FT DOMAIN 501 557 COILED COIL (POTENTIAL).  
FT DOMAIN 607 617 PRO-RICH.  
FT DOMAIN 988 1816 INTERACTION WITH ETO.  
FT DOMAIN 2055 2059 CORNR BOX OF ID1.  
FT DOMAIN 2263 2267 CORNR BOX OF ID2.  
FT DOMAIN 58 64 POLY-GLN.  
FT DOMAIN 593 603 POLY-ALA.  
FT DOMAIN 1032 1035 POLY-PRO.  
FT DOMAIN 1707 1712 POLY-ALA.  
FT DOMAIN 1952 1963 POLY-SER.  
FT CONFLICT 1014 1014 L -> V (IN REF. 2).  
FT CONFLICT 1508 1509 PP -> SS (IN REF. 2).  
FT CONFLICT 1561 1561 W -> R (IN REF. 2).  
FT CONFLICT 1567 1567 Q -> H (IN REF. 2).  
SQ SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

Query Match 100.0%; Score 12643; DB 1; Length 2440;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2440; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSSSGYPNQGAFSTEQSRYPHPSVQVTFPTNRHQEQFAVPDVRSSHLEVSAQSOLLQQQ 60  
DB 1 MSSSGYPNQGAFSTEQSRYPHPSVQVTFPTNRHQEQFAVPDVRSSHLEVSAQSOLLQQQ 60  
QY 61 QQQLRRRPSLLSEFFHSGDRPQERRTSYEPFHGPPSPVDHDSLESKRPRLEQVSDSHFQ 120  
DB 61 QQQLRRRPSLLSEFFHSGDRPQERRTSYEPFHGPPSPVDHDSLESKRPRLEQVSDSHFQ 120  
QY 121 RVSAVPLVHPLPEGLRASADAKDPAGKGHEAPSSIPGQPCGDDQNAPSLSKEE 180  
DB 121 RVSAVPLVHPLPEGLRASADAKDPAGKGHEAPSSIPGQPCGDDQNAPSLSKEE 180  
QY 181 LIQSDMRDREIAKVEQOILKKKQOOLEEAAKPEPEKVPSPVPVQKHRSIVQIY 240  
DB 181 LIQSDMRDREIAKVEQOILKKKQOOLEEAAKPEPEKVPSPVPVQKHRSIVQIY 240  
QY 241 DENRKAEEAHKIEFGLGPKVELPLYNQPSDTKYVYHENIKTNQVARKKLIFFKRRNHAR 300  
DB 241 DENRKAEEAHKIEFGLGPKVELPLYNQPSDTKYVYHENIKTNQVARKKLIFFKRRNHAR 300  
QY 301 KORQOKICQRYDQLEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360  
DB 301 KORQOKICQRYDQLEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360

DB 301 KORQOKICQRYDQLEAWKVKVDRIENNRKAKESKTREYIEKQFPEIRKQEQBERFQ 360  
QY 361 RVGORGAGLSATARSHEISEIIDGLSEQENNEKQRLQSVIPPMFDAQORVKFINM 420  
DB 361 RVGORGAGLSATARSHEISEIIDGLSEQENNEKQRLQSVIPPMFDAQORVKFINM 420  
QY 421 NGLMEDPMKVKYKDRQFMNVWTDHEKEIFKDKFIOHPKNFGLIASYLERKSPDCVLVYLL 480  
DB 421 NGLMEDPMKVKYKDRQFMNVWTDHEKEIFKDKFIOHPKNFGLIASYLERKSPDCVLVYLL 480  
QY 481 TKKNYKALVRRNYGRRGRNQIARPSQEEKVEEKEEDKAETKEEKKDEEKEDE 540  
DB 481 TKKNYKALVRRNYGRRGRNQIARPSQEEKVEEKEEDKAETKEEKKDEEKEDE 540  
QY 541 KEDSKENTKEDKTDGTAETETEEEOATPGCKTANSQGRBKGRITBSMTNEAASAAA 600  
DB 541 KEDSKENTKEDKTDGTAETETEEEOATPGCKTANSQGRBKGRITBSMTNEAASAAA 600  
QY 601 AAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNNAIAKMYGTKEA 660  
DB 601 AAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNNAIAKMYGTKEA 660  
QY 661 QCKNFYFNYKRRHLDNLQHQKTSRKPRERDVSCQESVASTVSAQEDIEASNEE 720  
DB 661 QCKNFYFNYKRRHLDNLQHQKTSRKPRERDVSCQESVASTVSAQEDIEASNEE 720  
QY 721 ENPEDSEVAVKPSEDSPENATSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDSV 780  
DB 721 ENPEDSEVAVKPSEDSPENATSRGNTPEPAVELEPTTETAPSTSPSLAVPSTKPAEDSV 780  
QY 781 ETQVNDISIAETAQMDVDQEQHSAEGSVCDPPATKADSDVDVEVRYPENHASKVEGDN 840  
DB 781 ETQVNDISIAETAQMDVDQEQHSAEGSVCDPPATKADSDVDVEVRYPENHASKVEGDN 840  
QY 841 TKERDLRASEKVEPRDELVVAQIINAQRPESQDNDSSATCSADEVDGEPERQRMFP 900  
DB 841 TKERDLRASEKVEPRDELVVAQIINAQRPESQDNDSSATCSADEVDGEPERQRMFP 900  
QY 901 MDSKPSLLNPTGSLVSSPLKPNLDLPOLQHRAAVTPMVSCPTCNIPICPTVSGVALY 960  
DB 901 MDSKPSLLNPTGSLVSSPLKPNLDLPOLQHRAAVTPMVSCPTCNIPICPTVSGVALY 960  
QY 961 QRHIKAMHESALLEEQRQEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLITNLP 1020  
DB 961 QRHIKAMHESALLEEQRQEQIDLECRSSTSPCGTSKSPNREWEVLQAPAPHLITNLP 1020  
QY 1021 GVRLLPTRPRPPPLIPSSKTTVASEKPSIFMGSGSISQGTPTGYLTSHNQASTQETPK 1080  
DB 1021 GVRLLPTRPRPPPLIPSSKTTVASEKPSIFMGSGSISQGTPTGYLTSHNQASTQETPK 1080  
QY 1081 PSVGSISLGLPRQESAKSATLPYIKOEFSRSONSQPEGLLVRAOHEGVVRCGTAGAIQ 1140  
DB 1081 PSVGSISLGLPRQESAKSATLPYIKOEFSRSONSQPEGLLVRAOHEGVVRCGTAGAIQ 1140  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVKGSI SRMPIDESSPEKG 1200  
DB 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVKGSI SRMPIDESSPEKG 1200  
QY 1201 REEAASGHVIEYEGKSGHILSYDNKNAREGTRSPRTAHEISLKRSTESVEGNTKQGM 1260  
DB 1201 REEAASGHVIEYEGKSGHILSYDNKNAREGTRSPRTAHEISLKRSTESVEGNTKQGM 1260  
QY 1261 RESPVSAPEGLICRALPRGSPHSDLKERTVLSGIMQGTPTATTESFEDGLKYPKOIKR 1320  
DB 1261 RESPVSAPEGLICRALPRGSPHSDLKERTVLSGIMQGTPTATTESFEDGLKYPKOIKR 1320  
QY 1321 ESPPIRAFEAGITKGPYDGIITIKEMGRSITHEIPRODILTQESRKTPVQVQSTRPIEG 1380  
DB 1321 ESPPIRAFEAGITKGPYDGIITIKEMGRSITHEIPRODILTQESRKTPVQVQSTRPIEG 1380  
QY 1381 SISOGTPIKFDNNSGQSAIKHNKSLITGSKLSRGMPPLIETVPEKNKVERGKYEDVKA 1440  
DB 1381 SISOGTPIKFDNNSGQSAIKHNKSLITGSKLSRGMPPLIETVPEKNKVERGKYEDVKA 1440











QY 1835 DYSKTESKHEAARLEENLRSAASVSEQQLEQKLEVEKRSVQCLXTSSAFFSGKQPQ 1894  
Db 1851 DYSKTESKHEAARLEENLRSAASVSEQQLEQKLEVEKRSVQCVCTSSALPSGKAQ 1910  
QY 1895 HSSVYSAGADKGGPPKRSYEEELTRGKTITITAAFDIVITITQIASDKDARGSGS 1954  
Db 1911 HASVYSAGADKGGPPKRSYEEELTRGKTITITAAFDIVITITQIASDKDARGSGS 1970  
QY 1955 SDSSSLSHRYETPSDAIEVISPASSAPPOEKIQTQPEVVKVANAQENDPTROYEGL 2014  
Db 1971 SDSSSLSHRYETASDAIEVISPASSAPPOEKIQTQPEVVKVANAQENESTROYEGL 2030  
QY 2015 HHYRQOESPQQO - LPPSSQAEGMGQVPRTHLITLADHICQIITQDFARNQVSSQT 2072  
Db 2031 HHYRQOESPQQOPLPPSSQSEGMQVPRTHLITLADHICQIITQDFARNQV - - - 2086  
QY 2073 PQOPTSTFONSPSALVSTPYRTKTSNRYSPESAQSVHHQRPGRSVSPENILVDRSGSR 2132  
Db 2087 PSQASTSTFOTSPSALSSTPVRTKTSNRYSPESQSTQVLHPRPGPRVSPENILVDRSGSR 2146  
QY 2133 PGKSPERSHVSPEYEPISPPQVPPVHVHEKQDLSLLSQRGAEPABQNDARSPGISYLP 2192  
Db 2147 PGKSPERSHVSPEYEPISPPQVPAVHEKQDLSLLSQRGVDPABQNDARSPGISYLP 2206  
QY 2193 SFTKLENTSPMVSKKQEIFRKLNSGGGSDMAAOPGTEIFNLPAVTTSGSVSSRGH 2252  
Db 2207 SFTKLENTSPMVSKKQEIFRKLNSGGGSDMAAOPGTEIFNLPAVTTSGAVSSRSH 2266  
QY 2253 SPADPASNLGLEDIIRKALMSFDDKVEDHGVVMSQPMGVPGTANTSVTSGEFTREEG 2312  
Db 2267 SPADPASNLGLEDIIRKALMSFDDKVEDHGVVMSHPYIMPGSASTSVTSSEARRDEG 2326  
QY 2313 DPSHSGGVCKPKLKSNSRKSPIPGQYLGTPERPSSVSVHSEGDYHQRTPGWAE 2372  
Db 2327 EPSHA - GVCPRKLINKSNRKSPIPGQYLGTPERPSSVSVHSEGDYHQRTPGWAE 2385  
QY 2373 DRPSSTGSTQPPYNPLTWMLSSSTPTPIACAPSAVNAQAHPHQNRWEREPAPLLSAQY 2432  
Db 2386 DRPSSTGSTQPPYNPLTWMLSSSTPTPIACAPSAITQAHPHQNRWEREPAPLLSAQY 2445  
QY 2433 ETLSDSD 2440  
Db 2446 ETLSDSD 2453

RESULT 3  
NCR2\_HUMAN  
ID NCR2\_HUMAN STANDARD; PRT; 2517 AA.  
AC Q9Y500; Q9Y500; Q13354; Q00613; O15416;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF  
DE RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (SMRTE) (THYROID-,  
DE RETINOIC-ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-  
DE ASSOCIATING FACTOR) (TRAC) (CTG26).  
GN NCR2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC TISSUE-Pituitary;  
RX MEDLINE=93178941; PubMed=10077563;  
RA Orntlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;  
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC TISSUE-Cervix adenocarcinoma;  
RX MEDLINE=99199215; PubMed=10097068;  
RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;

"SMRTE, a silencing mediator for retinoid and thyroid hormone  
receptors-extended isoform that is more related to the nuclear  
receptor corepressor.";  
Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
RL [3]  
RP SEQUENCE OF 1023-2517 FROM N.A.  
RN TISSUE-Cervix adenocarcinoma;  
RX MEDLINE=96008552; PubMed=7566127;  
RA Chen J.D., Evans R.M.;  
RT "A transcriptional co-repressor that interacts with nuclear hormone  
receptors.";  
Nature 377:454-457(1995).  
RL [4]  
RP SEQUENCE FROM N.A. (ISOFORM TRAC-1).  
RN TISSUE-Fetal liver;  
RX MEDLINE=96408715; PubMed=8813722;  
RA Sande S., Privalsky M.L.;  
RT "Identification of TRACs (T3 receptor-associating cofactors), a family  
of cofactors that associate with, and modulate the activity of,  
nuclear hormone receptors.";  
Mol. Endocrinol. 10:813-825(1996).  
RL [5]  
RP SEQUENCE OF 428-613 FROM N.A.  
RN TISSUE-Brain cortex;  
RX MEDLINE=97369492; PubMed=9225980;  
RA Margolis T.S., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,  
Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
RT "CDNAs with long CAG trinucleotide repeats from human brain.";  
Hum. Genet. 100:114-122(1997).  
RL [6]  
RP NUCLEAR RECEPTOR-ASSOCIATING COFACTOR (TRAC-1) CONTAINS ONLY  
THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN  
ANTI-PRESSOR.  
CC [7] TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE  
DETECTED IN LUNG, SPLEEN AND BRAIN.  
CC [8] INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.  
CC [9] DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
DOMAINS (ID1 AND ID2).  
CC [10] DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND  
SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES  
FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
SPECIFICITY.  
CC [11] SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).  
CC [12] SIMILARITY: CONTAINS 2 CORN BOX.  
CC [13] SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF113003; AAD20946.1; -  
DR EMBL; AF125672; AAD22973.1; -  
DR EMBL; U37146; AAC50236.1; -  
DR EMBL; S83390; AAB50847.1; -  
DR EMBL; U80750; AAB91446.1; -  
DR MIM; 600848; -



DR InterPro; IPR001005; ..  
DR pfam; PF00249; myb\_DNA-binding; 2.  
DR PROSITE; PS0090; MYB\_3; 1.  
KW Nuclear protein; transcription regulation; DNA-binding; Repressor;  
FT Coiled coil; Alternative splicing.  
FT DOMAIN 174 215  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 254 312  
FT INTERACTION WITH SIN3A/B (BY SIMILARITY).  
FT DNAS\_BIND 429 474  
FT SANTI-A (POTENTIAL).  
FT DNAS\_BIND 612 657  
FT SANTI-B (POTENTIAL).  
FT DOMAIN 522 561  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 778 820  
FT PRO-RICH.  
FT DOMAIN 2139 2143  
FT CORNER BOX OF ID1.  
FT DOMAIN 2342 2346  
FT CORNER BOX OF ID2.  
FT DOMAIN 494 510  
FT POLY-GLN.  
FT DOMAIN 682 685  
FT POLY-LYS.  
FT DOMAIN 994 1002  
FT POLY-PRO.  
FT DOMAIN 1384 1389  
FT POLY-PRO.  
FT DOMAIN 1842 1846  
FT POLY-GLY.  
FT DOMAIN 2479 2482  
FT POLY-PRO.  
FT VARSPLIC 1 1702  
FT MISSING (IN ISOFORM TRAC-1).  
FT VARSPLIC 2353 2398  
FT MISSING (IN ISOFORM TRAC-1).  
FT CONFLICT 7 7  
FT L -> P (IN REF. 2).  
FT CONFLICT 295 295  
FT K -> E (IN REF. 2).  
FT CONFLICT 309 309  
FT L -> W (IN REF. 2).  
FT CONFLICT 352 352  
FT MISSING (IN REF. 2).  
FT CONFLICT 365 365  
FT A -> P (IN REF. 2).  
FT CONFLICT 612 613  
FT SS -> EF (IN REF. 5).  
FT CONFLICT 711 711  
FT S -> T (IN REF. 2).  
FT CONFLICT 724 740  
FT MISSING (IN REF. 2).  
FT CONFLICT 787 796  
FT RTRRAPLEP -> PEDIPAPTES (IN REF. 2).  
FT CONFLICT 804 804  
FT G -> L (IN REF. 2).  
FT CONFLICT 814 814  
FT S -> F (IN REF. 2).  
FT CONFLICT 817 817  
FT A -> S (IN REF. 2).  
FT CONFLICT 889 889  
FT G -> R (IN REF. 2).  
FT CONFLICT 1023 1030  
FT SRSAPPA -> MEAWDAHP (IN REF. 3).  
FT CONFLICT 1034 1034  
FT A -> AEKVPFPA (IN REF. 2).  
FT CONFLICT 1894 1894  
FT K -> T (IN REF. 4).  
FT CONFLICT 2494 2494  
FT P -> A (IN REF. 4).  
SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;

Query Match 32.8%; Score 4147.5; DB 1; Length 2517;  
Best Local Similarity 40.7%; Pred. No. 1.3e-146;  
Matches 1099; Conservative 341; Mismatches 785; Indels 475; Gaps 108;

Qy 17 QSRPPHSHVQTFNTRHQEFVDPDYRSHSHLEYSQASQLLQOQQOQLRRPSSLSEFH 76  
Db 16 EPRYPHSHVQTFNTRHQEFVDPDYRSHSHLEYSQASQLLQOQQOQLRRPSSLSEFH 73  
Qy 77 PGSDRPQE--RRTSYEPHPGSPVDHDSLSKSRPLRQVSDSHFQVSAALPLVHPLP 134  
Db 74 PGNERSQELHLPESHVLPGLKSEMEFIESKRLLELPD-----PLLRSP 122  
Qy 135 ---EGLRA-SADAKKDPAGKHEAPSSPSIGQCGDDQNASPKLSKEELIQSDRVDR 190  
Db 123 LLATQAPAGSEDLTKDRSLTKLE--PVSPGPPHPTDPELELVPLRSLKEELIQSDRVDR 181  
Qy 191 ETAKVEQILKKKQOQLEEEAKPPEKPVSPVPPVEQKHSRSTVQIYYDENRKAEEA 250  
Db 182 EITMVEQILKKKQOQLEEEAKPPEKPVSPVPPVLESKHSRLVQIYYDENRKAEEA 241  
Qy 251 HKIFEGLPKVELPLYNQPSDTKVYHENIKTNQVMRKLILFFKRRNHARKQEQKICOR 310  
Db 242 HRILGGLPQVELPLYNQPSDTQYHENIKINQAMRKLILYFKRRNHARKQEQKICOR 301  
Qy 311 YDQLEMEAKKVDRIENNRKAKESKTRYEYKQFPIRQKQOQERFQ-RVGORGAGL 369  
Db 302 YDQLEMEAKKVDRIENNRKAKESKTRYEYKQFPIRQKQOQERFQ-RVGORGAGL 361  
Qy 370 SATATRSHEITSEIIDGLSEQENNEKQRLSVIPPMFDEAQRVKFINNGLMEEDPMK 429  
Db 362 SMSAARSEHVEIIDGLSEQENNEKQRLSVIPPMFDEAQRVKFINNGLMEEDPMK 421

Qy 430 VYKROPMNVWTDHEKEIFKDKFIOHPKNFGLIASYLKRSKSPDCVLYYLYTKKNENYKA 489  
Db 422 VYKROPMNVWTDHEKEIFKDKFIOHPKNFGLIASYLKRSKSPDCVLYYLYTKKNENYKA 481  
Qy 490 LVRRNYGRRNRNQIARPSQEEKVEEKE--DKAEKTEKKEEKKDEEKKDEKDSKEN 547  
Db 482 LVRRSY--RRRKSQOQQOQQOQQOQQOQQMPRSSQEEKDEKEKEKEEKEEKEPEVEN 540  
Qy 548 TKE---KOKIDGTA-EETEERQATPRGRKTKANSOGRKRGKGRITRMTNEAAAAA 603  
Db 541 DKEDLLKEKTDGTDGDNDEKAVASKRKTANSOGRKRGKGRITRMTNEAAAAA 600  
Qy 604 TEEPPPLPPPEPTEPSTETSWTEEMEVAKGLVEHGRNNAIAAKMVKTSKSEACK 663  
Db 601 S-----AELASMEINESSRTEEMETAKGLLEHGRNNSAIARMVSKTVSOCK 650  
Qy 664 NFYFNYKRRHNDLLOQHKQKTRKPREEDVDSQESVASTVSA---QEDEDIAS-- 717  
Db 651 NFYFNYKRRQNLDELQHLKMEKERNARRKKKAPAAASEEAAFPVVEDEMEASGV 710  
Qy 718 --NEENPEDSE-----VEAVK----- 732  
Db 711 SGNEEMVEEAEALHASGNEVPRGECGSPATVNNSSDTEPSPHTEAAKDTQNGPKPP 770  
Qy 733 -----PSEDSPENATSRGNTPEPAVELE---PTTETAPSTSPSLAVSTKPADES 779  
Db 771 ATLGADGPPGPPPPPTPRRTSRAPTEPTPASEATGAPTPPAP--PSPS-APPVVPKEEKE 828  
Qy 780 VETQVNDISAEATQMDVDQOEHSAEESGVCDDPPPATK--ADSDVDVVRVVRPENHASVE 837  
Db 829 EETAAAPPV-----EBGERQKPPAAELAVDTGKAEEPVKSECTEAE 871  
Qy 838 GDNTERDLRA-----SEKVEPRDEDLVVAQINAORPEQSDNDSSATCSADE- 887  
Db 872 EGPAKGDAEAAEAETAGALKAEKKEGGSRATTAKSSGA----PQ-DSDSATCSADEV 926  
Qy 888 -DVGEPEPORRFPMDSKPSLLNPTGSLV--SSPLKPNPLDLQLOHRAAVIPPMVSCPT 945  
Db 927 DEAGGDKNRLSP---RPSLLTPTGDDPRANASQK--PLDLKQLKQRAAIPPI----- 976  
Qy 946 CNIPIGTPVSGYALYORHIKAMHESALLE-----EORQORQOIDEKRSSTSP 993  
Db 977 -----QVTKVHEPPREDAAATKPAAPPQPPQNLQPPQNLQPPQNLQPPQNLQPPQ 1019  
Qy 994 CGTSKSP---NRE-----W-----EVLOPAPH-----QLITNLP 1019  
Db 1020 RGRSRSPAPPADKFAEAAEAKLPDPPCWTSGLPFPVPPPREVIKASHPADPPSAFAP 1079  
Qy 1020 EGVRLP-----TTTRTPPPPLIPSSKTTVAASEK-PSFI--MGSSISQGTPTGTYLTSNQ 1071  
Db 1080 PGHPLPLGLHDTARVLP RPPTISNPPPLISSAKHPVLERQIGAISQG-----MSVQLH 1134  
Qy 1072 ASYQETPKPSVGSISGLPRQOESAKSATLPYIKQEFSPRSQNSQPEGLLV-RAQHGH 1130  
Db 1135 VPYSEHAKAP--VGPVTMGLPLPMDPKLAPESGVKQEQLSRPGQAGPPESLGVPTAQAS 1193  
Qy 1131 VVRGTA-GAIOEGSTTRCTPTSKISVESIPSLRGSTTGTPTALPQTGPTALVKGSTSR 1189  
Db 1194 VLRTGALGVPGGSTIGIPSTRVPSDAITRGSTHGTGA-----DVLKGTITR 1245  
Qy 1190 MPIDSSP--EKGREAAKSHVIEYEGKSHLSYDNIKNA-----REGTRSPRTAHS 1242  
Db 1246 IIGEDSPRLDRGREDLSLPKGVHIEYEGKSHLSYDNIKNA-----REGTRSPRTAHS 1305  
Qy 1243 L-KKSYVESVGNIKOGMSMRSPVSALEGLICRALP--RGSPHSDLKERTVLSGIMQG 1299  
Db 1306 APKRTYDMGVRGVRATIS-----SASIEGLMGRAPPERHSPH-HLKEQHHSIGSTQG 1358  
Qy 1300 TPRTATESFEDGL-KYPKQIKRES-----PPIRAFEGAI-----TKGKQ-YDG-ITFI 1344  
Db 1359 IPRSYVEAQEDYLRBAKLLKREGTPPPPPSRDLTEYKTAQALGPLKLAHAELGATV 1418  
Qy 1345 KEMGRSITHEIPRODILTQESRKTPEVQSTRPIIEGSIQGTPIKFDNN-SQSAIKHNV 1403











Db 1127 ---LGAISQOQMSVOLRVPHS-----EHAKAPMGPIITMGLPLAVDPKKLGT----- 1169  
Qy 1110 FSPRSQNSQPEGLLVRAQHEGVVGTAGTAQEGSITRTGTPTSKISVESIPSLRGSITQGT 1169  
Db 1170 ---ALGATSQSGSITKGLPSTRAA---DGPSYRGSITHT 1202  
Qy 1170 PALPQTGIPTEALVKGSIIRMPIDSSP---EKGREEAASGHVYIEKSGHILSYDNKN 1227  
Db 1203 PA-----DVLVKGITISRVGDSPSRDLRAREDTLPKHVYIEKSGHVLSEYEGMS 1254  
Qy 1228 A-----REGTRSPRTAHEISL-KESYSEVGENIKQGMSPRESVPASPLEGLICRALP-RG 1280  
Db 1255 VSQCKEDGRSSGPPHETRAKPTIDMEGRVGRITV-----SASIEGLMGRAPTEQH 1308  
Qy 1281 SPHDLKERTVLGSIIMQGTTPRATESFDEL-KYPOIKRES-----PPIRAFEGAITK 1334  
Db 1309 SPH--LKEQHIIIRSGITQIGIPRSVVEAQEDYLRREKLLKREGTPPPPPPRDLTEY-K 1365  
Qy 1335 GKPYDG-----TTTIKEMGRSHEIPRQDILTOESRKTTEVVOSTRPIIEGS 1381  
Db 1366 PRPLDPLGLKLPKTHEGVVATVKEAGRSIHEIPREEL-----RRTPELPLAPRLKEGS 1420  
Qy 1382 ISQGTPIKEDNNSQSAI-KHNKSLITGSKLSRGMPLEIVPENIKVVERGKYEDVKA 1440  
Db 1421 ITQGTPLKYGAPSTGKHKHVSRIIGSGRPPPALHPLDIMAD-ARALERACYE----- 1475  
Qy 1441 GETVRSRHTSVSSGSPVLRS---TLHEAKAQLSPGIDYDTSARRTPVSYONTMSRGP 1497  
Db 1476 -ESLKSRSSTSSGAGSITRCAPVVVPELGPQSPITIEDHGA-----PPTSHLPGSP 1529  
Qy 1498 MNNTSDVTIPP-----NKSTNHERKSTLTPTQRESIPAKSPVPGVDVPSVSH--SPFDPHH 1551  
Db 1530 VTTREPTPLRQEGSLLSKASQDRKLTSTPRE-----TAKSPHSTVPEHHPHPSYEHLL 1585  
Qy 1552 RGSTAGEVYWSHLPTQLDP-AMPHRALDPAAAYLQORQLSPPTPGYPSQYLY----- 1604  
Db 1586 RGVTVGLRGHILAFDPTISIPRGIPLEAAAAAYILPRHLAPSPTPHLYPPYLIRGYP 1645  
Qy 1605 ---AMENTROTILNDYITSOQMVLN-----RPDVARGLSPREQPLGLPYPA-TRGIID 1654  
Db 1646 DTALEN-ROTIINDYITSOQHNNHNAASMAQADMLRGLSPRESSIALNYAAGPRGIID 1704  
Qy 1655 LTNMPP-TILVPHPGGTSTPMDRITVPGTITPPRPYNSASMSFGHPHLA-AAASA 1712  
Db 1705 LSQVPHPLVLPVPPGTGPATAIDRLAYLPTAPPFSSK-HSSSPLSPGPGPHLAKPTATS 1763  
Qy 1713 ER 1763  
Db 1764 SSER 1822  
Qy 1764 QETMLQORPSVFOQTNGTSTVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADALAA 1823  
Db 1823 QDA-LQORPSVHLNHTSMKGVVTSVEP-----GTPTV-----LRW 1855  
Qy 1824 LVDAASAPQMDVSKTESHEAARLEENLRSAVSEQOQLBQKTEV---EKRVSQC 1880  
Db 1856 ARSTSTSPVPAATFPATH--CPGLGTLGVVYPTLMPEVLLPKETSRVARPERARVDA 1913  
Qy 1881 LYTSSAFPSGKQPHSSVYSEACKDGPPIKRYEELTRGTITTAANFIDVITRQ 1940  
Db 1914 GHAFLTTPGGR-EPASPSKSSSEPRSLAPFSSH-----TAIART 1952  
Qy 1941 IASDKARERGSQSSDSSLSLHRYTTPS-----DAIEVIS 1977  
Db 1953 PAKNLAPHASDPDPATPSADLHREKTQKPSIQEILRLSLYHSGAGYSPGCVPEIS 2012  
Qy 1978 PASSPAPPQEKLOTQYQEVKVAQNAENDPTQYEGP-----LHRYP---QOESPS 2025  
Db 2013 PVSSPSLTHDKGLSKPLEEKSHLEGELRHKQPGMKLSAEAAHLPHRLPSPQSPSS 2072  
Qy 2026 PQOQLPSSQAEAGMQVPRHRLITLADHICQITQDFARNQVSSQTPQO---PPTSTFQ 2082  
Db 2073 PLIOTAPGIRGH-----QRVVTLAQHISEVITQDTRHH-----PQOLSGPLAPLY 2119

Qy 2083 NSPSALYSTPV---RTKTSNRYSPESQAQSVHHQRPGRSVSPENLVKSRGSRPGKSPER 2139  
Db 2120 SFGCA--SCPVLDLRRPSPDLPLPPD-----HGTTP-ARGSPH-----SEGKRSPEPSK 2166  
Qy 2140 SHV---SSEPYEPISPQO--VPVYHEKQDSLILLLSQSGABEAEQNRDARSFGSILYLSF 2194  
Db 2167 TSVLGSSDATEIPVSPPEGMTEPGHARSTAYPLL-YRDEQGEPRMGLESFGNTSQPPTF 2225  
Qy 2195 FTKL-ENTSPMVSKKQEIFRKLNSGGGSDMAAOPGTGIFNLPAVTTSGSVSSRGHS 2253  
Db 2226 FSKLTESNANVSKKQKQKINLKNHNEPEYINIGPTEIFNPAITAGLMTCSRQA 2285  
Qy 2254 FADPAS-NLGLIEDIIRKALMGSDFKVEDHGVMSQPMGV---VPGTANTSV-----VT 2303  
Db 2286 VQEHASTNMGLEAIIRKALMGKYQWEB-----PPPLGANAFNPLNASILPAAAMPIT 2339  
Qy 2304 SGETRREEDGSPHSGVCKPKLISKSNRSKSPKIPQCGYLCTGTERPSSVSVHSEGYH 2363  
Db 2340 TADGRSDHALTSPGGG--KAKVSGRPSRRKAKSPAP--GLASGDRPPSVSVHSEGCN 2395  
Qy 2364 ROTP--GWAWEDRPSSTGTOFPYNPLTMRM---LSSTPPTPIACAPSAYNOAAPHQON 2417  
Db 2396 RRTPLTNVWEDRPSAGSTFPFNPMLRLQAGVMSPPPPGLAAGSPL--AGPH--- 2450  
Qy 2418 RIWEREPAPLLSAQYETLSDSD 2439  
Db 2451 HAWDEEPKPLCSQYETLSDSE 2472

RESULT 5  
NCRL\_RAT  
ID NCRL\_RAT STANDARD; PRT; 533 AA.  
AC O9KUB5: 070463;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR) (FRAGMENT).  
GN NCOR1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP TISSUE=Brain;  
RC MEDLINE=99371771; PubMed=10441327;  
RA Boutell J.M., Thomas P., Neal J.W., Weston V.J., Duce J., Harper P.S., Jones A.L.;  
RT "Aberrant interactions of transcriptional repressor proteins with the  
RL Hum. Mol. Genet. 8:1647-1655(1999).  
RN [2]  
RP SEQUENCE OF 476-528 FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=99421707; PubMed=10491148;  
RA Schuler M.J., Buehler S., Pette D.;  
RT "Effects of contractile activity and hypothyroidism on nuclear hormone  
receptor mRNA isoforms in rat skeletal muscle.";  
RL Eur. J. Biochem. 264:982-988(1999).  
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
ABSENCE OF LIGAND (BY SIMILARITY).  
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
DOMAINS (ID1 AND ID2).



CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
CC SPECIFICITY.  
CC  
CC -1- SIMILARITY: CONTAINS 2 CORNR BOX.  
CC  
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF124821; AAD32566.1; -  
CC EMBL: AF059311; AAC14567.1; -  
CC PROSITE: PS00090; MYB.3; PARTIAL.  
CC Nuclear protein; Transcription regulation; DNA-binding; Repressor.  
KW NON\_TER 1 1  
FT DOMAIN 48 59 POLY-SER.  
FT DOMAIN 153 157 CORNR BOX OF ID1.  
FT DOMAIN 357 361 CORNR BOX OF ID2.  
FT CONFLICT 484 484 R -> W (IN REF. 2).  
FT CONFLICT 497 497 A -> V (IN REF. 2).  
SO SEQUENCE 533 AA; 57794 MW; 7DF60F8228227EC2 CRC64;

Query Match 19.7%; Score 2493.5; DB 1; Length 533;  
Best Local Similarity 90.08; Pred. No. 4e-86;  
Matches 484; Conservative 21; Mismatches 26; Indels 7; Gaps 3;

Qy 1905 KDKGPKPKRYEELRTRGKTTITAAFNIDVITQIASDKDAREGSQSSSSLSH 1964  
Db 1 KDKGPKPKRYEELRTRGKTTITAAFNIDVITQIASDKDAREGSQSSSSLSH 60  
Qy 1965 RYETPSDAIEVISPASSAPPOEKLQYTOPEVVKANAENDPTQYEGPLHYRPOQESP 2024  
Db 61 RYETPSDAIEVISPASSAPPOEKLQYTOPEVVKANAENDPTQYEGPLHYRPOQESP 120  
Qy 2025 SPOQOQ--LPPSSQAGMGQVPRHLITLADHICQIITQDFARNQVSSQTPQPPSTTFQ 2082  
Db 121 SPOQOQPLPPSSQAGMGQVPRHLITLADHICQIITQDFARNQV--PSQPSTSTFQ 176  
Qy 2083 NSPSALYSTPVRTTSNRYSPESQAQSVHHPGCRSPENLVKSGRSPKSPERSHV 2142  
Db 177 TSPSALSSTPVRTKPSRYSPESQSVQLVHPRGPRVSPENLVKSGRSPKSPERSHI 236  
Qy 2143 SSEPEYTPSPQVPPVHVEKQDLSLLSQGAEPAEQNDARSPGISYLPSPFTKLENTS 2202  
Db 237 PSEPEYTPSPQVPPVHVEKQDLSLLSQGMDEQNDARSPGISYLPFTFKLENTS 296  
Qy 2203 PMVSKKQEIFRKLNSGGGSDMAAQAQPGTEIFNLPAVTTSGVSRGSHSFADPASNLG 2262  
Db 297 PMVSKKQEIFRKLNSGGGSDMAAQAQPGTEIFNLPAVTTSGAVSRSHSFADPASNLG 356  
Qy 2263 LEDIIRKALMGSDKVEDHGVMSQWGVVPGTANTSVTSGTTRREGDPPSPHSGVC 2322  
Db 357 LEDIIRKALMGSDKVEDHGVMSQWGVVPGTANTSVTSGTTRREGDPPSPHSGVC 415  
Qy 2323 KPKLINSNRKSPKIPGQCYLGTERTPSSVSVSHSGDYHROTPGAWEDRPSSTGQ 2382  
Db 416 KPKLINSNRKSPKIPGQCYLGTERTPSSVSVSHSGDYHROTPGAWEDRPSSTGQ 475  
Qy 2383 FPNPLTMRMLSSPTPTPIACAPSAVNAQAAPHQONRIWEREPALLSAQYETLSDSD 2440  
Db 476 FPNPLTMRMLSSPTPTPIACAPSAITQAAPHQOSRIWEREPALLSAQYETLSDSD 533

RESULT 6  
MAPB\_HUMAN

ID MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
AC P46821;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MICROBULE-ASSOCIATED PROTEIN 1B [CONTAINS: MAP1 LIGHT CHAIN LC1].  
GN MAP1B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95104835; PubMed=7806212;  
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;  
RT Cloning of human microtubule-associated protein 1B and the  
RL Genomics 22:273-280(1994).  
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAP1B IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -1- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE  
CC KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER  
CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES  
CC BOTH IN VITRO AND IN VIVO.  
CC -1- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
CC BOTH MAP1A AND MAP1B.  
CC -1- SIMILARITY: TO NEURAXIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: L06237; AAA18904.1; -  
CC MIM: 157129; -  
CC InterPro: IPR000102; -  
CC Pfam: PF00414; MAP1B\_NEURAXIN; 10.  
CC PROSITE: PS00230; MAP1B\_NEURAXIN; 6.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.  
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
FT KKEE AND KKEI/V REPEATS).  
FT DOMAIN 1869 2074 12 X 17 AA TANDEM REPEATS.  
FT REPEAT 1869 1885 1.  
FT REPEAT 1886 1902 2.  
FT REPEAT 1903 1919 3.  
FT REPEAT 1920 1936 4.  
FT REPEAT 1937 1953 5.  
FT REPEAT 1954 1970 6.  
FT REPEAT 1971 1987 7.  
FT REPEAT 1988 2004 8.  
FT REPEAT 2005 2021 9.  
FT REPEAT 2022 2038 10.  
FT REPEAT 2039 2055 11.  
FT REPEAT 2056 2072 12.  
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 3.8%; Score 480; DB 1; Length 2468;  
Best Local Similarity 18.7%; Pred. No. 8.8e-11;  
Matches 492; Conservative 375; Mismatches 1009; Indels 762; Gaps 113;

Qy 71 LLSEFHFGSDRPOQRRTSYEPFHFGSPVDHDSLESKRPRLEQVSDSHFORV---SAAVL 127







```
QY 2046 HRLI-----TLAD-----HICQIITQDFARNQSSVSTPQPPTSTF 2081
D 1981 RRLDDISNGYDDSDGHTLGDPSYSYETETKITSFPESEGSYVETSKTRTPDSTY 2040
QY 2082 -ONSFSALVSTPVRK-----TSNRYSPEQAQSV-----HHORPSGRVSPE 2122
D 2041 CYETAKEITRTPQASTSYETSIDLCTAETKKSPEARQDVCLVSSCEYKHPKTELSPS 2100
QY 2123 NLVDKSRGSRGKSPERSHVSSEPELSPQPVVWHEKQDLSLLLSORGAEPAE----- 2177
D 2101 FI-----NPNLEWFASEEPEESEK-----LTQSGGAPPPPGKQ 2137
QY 2178 --QRNDARSPGISYLPFFFTKLENTSMWKSKEIFRKLNSGGSDMAAAQPGTEI 2235
D 2138 QGROCDPTPTSVS--ESAPQOTSDVPETECPISITADANIDESEIPIPTDKVTY 2195
QY 2236 PNL-----PAVTTSGSVSRGH---SFADPASNLGLDIIIRKALMGFDDKVEDHGVMSQP 2289
D 2196 KMDPPAPVQDRSPR-HPDVSMVDPEA-LAIBQNLGKALKDKLKEKTK----- 2246
QY 2290 MGVPVGTANTSVVTSGETRREGDPSHGGVCRPKLLSKNSRSK--SPIPGGYLGT 2347
D 2247 ---KPGTKTS---SSPVKKSDDGSKPLAASP-KPAGLKESDVKSVRSVAPSKKESVEKA 2299
QY 2348 ERPSVSSVHS-----EGDYHRTQPCWAWEDRPSSTGSGTOFFVNPILTMRLSSTPTPI 2401
D 2300 AKPTTPEVKAARGEKDKETKNAANASAKSAXT-ATAGPTGTYTKSSAVPGPLPV 2356

RESULT 7
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.
AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=92355629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA.";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC -!- CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC -!- WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC -!- APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC -!- THEIR MORPHOLOGY.
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC -!- FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC -!- FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC -!- BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)

or send an email to license@isb-sib.ch.
CC -----
EMBL: M83196; AAB48069.1; -
DR PIR: A43359; A43359.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN 22465 2774 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
FT REPEAT 336 338 1.
FT REPEAT 415 417 2.
FT REPEAT 420 422 3.
FT REPEAT 424 426 4.
FT REPEAT 427 429 5.
FT REPEAT 431 433 6.
FT REPEAT 436 438 7.
FT REPEAT 440 442 8.
FT REPEAT 444 446 9.
FT REPEAT 449 451 10.
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299536 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 3.7%; Score 465.5; DB 1; Length 2774;
Best Local Similarity 19.1%; Pred. No. 3.5e-10;
Matches 570; Conservative 350; Mismatches 1012; Indels 1053; Gaps 140;

QY 96 PSPVDHDSLEKRPRLQVSDSHFQVRSAAVLPLVHPLPEGLRASADAKDPAGGKHEA 155
D 18 PSPFD-----LLEPPTSGGFLKLSK---PCYIFP-----GGRGDS 50
QY 156 PSSPISGQCGDDQNASPSKLSKEELQSDMRVDRE-IAKVEQQILK-----LKKKQOOL 209
D 51 ALFAVNGFNILVD--GGDRKSCFWKLVHRLDRISVLLTHIGADNLPGINGLLQKVAEL 109
QY 210 EEEAAK-----PPE-----PEKVPSPPPVEQKHRSIVQIILYDENRKAEE 248
D 110 EEEQSSSSYSDMWKNLISPELVGVFVFNVDKRLPDAARAKRSI-----E 157
QY 249 EA-----HKIPEGLGPKVELPLYNQPSDT---KVYH-----ENIKTNVMRKKLILF 292
D 158 EACITLQH--LNLRLGIAOE-PLYRVVSNITIEPLTLFHKMGVGRDLMYVLNPNVKSQEMQF 214
QY 293 FKRR---NHARKQ-----REQICQRYDQLEMA--W-----EKKVDRI---ENNPRR 331
D 215 LMQKWAGNSKAKTGIVLANGKEAELSVPLYTSITLVVWLPANPTEKIVRVLPFGNAPQN 274
QY 332 KAKE--SKTRYEYKQPP-----EIRKQEQOERFORVQORGAGLS 370
D 275 KILEGLEKRLHDLRYPVATQKDLAAGAVPANLKPSPKIKHRADSKESLKAAPTAV--- 331
QY 371 ATARSHEISEIIDGLSEQEN-----NEQMORQLSVIPPMFMFAEQRRVFINMNG 422
D 332 SKLAKREEVLEE---GAKEARSELAKELAKTEKKAKESEKPP-----EKPSKSERVG 382
QY 423 LMEDPMKYKDRQFMVWNTDHEKIEFKDKFTQHPKFNGLIASYLERKSVPCVLYYYLTK 482
D 383 ESSALKAERK-----LIKDK----- 399
QY 483 KNEYKALVRRNYKRRGRNQOIARPSQEEKVEKEEDAEKTEKKEEKEDEKDEKE 542
D 400 -----AGKKH-LKEKISK-LEEKKDKKEKKEKELKKEGKEEKEDAKK 445
QY 543 DSK-ENTKEKDKIDGTAETEERQATPRGKT---ANSOGR---RKGRITR---SMTNE 592
D 446 DEKRDKTKPEVK-----KLSKPDLPKPTFEVRYTKLYKAKAPGRVVDKRAARGEKELSS 501
QY 593 AAAASAAAAATEPPPPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLV----- 641
D 502 PRTPPAQKGA--PPAAVSGHRELALSSPEDLTQ-DFEELKREERGLLABQDRTGLGEK 557
QY 642 -----EHGRNWAAI--AKMVGTYSQAQCKNFYNTKRRHNLNLLOOHQKQTSRPR- 691
D 558 PLPADATEQGHPSAAIQVOTPSGPLEGE-----HVEREKEVVPDPSGDKGSTNRGPD 611
```



Qy 692 -----EEDVSOCSVASTVSAQEDIEDIASNEE--ENPDSRVEAVKPSDSPE 739  
Db 612 GAVEKEKETWEERKORAEALGPENTAARESEAEVREDVIEKAELEMEETHESDEGE 671  
Qy 740 NATSRG-----NTEPAVELEPTT-----ETAPSTSPSLAVS----- 771  
Db 672 ETKAESFYKHTQKALKASPKSREALGRDLGFOGKAPKETASFLS--SLATPAGATEHV 730  
Qy 772 -----TKPAEDESVEQVNDISIAETAE-----QMDVDOQEHSAEBSVCDDPP---AT 817  
Db 731 SYIOETIPGVSETEQITSDIEIHEDDERPAPRFTPTSYDLGPEG-----PGPFRAEQ 786  
Qy 818 KADSV-----DVEYRVP-----ENHAKVEGDNTKRDLDRASEKVEP 855  
Db 787 AADSAPASSSKTYGAPETELTYPPNMVAAPLAEHEHSSA-----TSITECDKLSSRATS 842  
Qy 856 RDELVVAQINARPE-----PQSDNDS-----SATCSADED 888  
Db 843 VAEDQSVASLTAPQTEETGKSSLLDVTISIPSSRTEATQGLDYPVSAGTISPTSSLEED 902  
Qy 889 -----VGEPE-----RORMRPMDSKPSLLNPTGSLILVSSP 919  
Db 903 KGFKSPCEDFSVTGSEKKGETVGRGLSGEKAVGKEKYVVTSEK--LSQYAAVFGAP 960  
Qy 920 ---LKNPPLDLPOLQHRA-----AVIPPMVSCPTPCNIPIGTPVSGYALYORHIKAM 967  
Db 961 GHTLPPGEPALGEVEERCLSPDDSTVKMASPP-----PSGPPSAHTPPF----- 1004  
Qy 968 HESALLEQRORQOIDLCHRSSTSPCGTSKSPNREVLQAPAHQLITNLPEGVRPLPT 1027  
Db 1005 HQSPVEDKSEPRDFQEDSWGTEKNSP--GVSKEDSEE--QTVKAPGPEEGTSE--EGKGPPTR 1060  
Qy 1028 RPTRPPPP-----LIPSSKTTVASEKPSFIMGSIQGT--PGYLTSHNOAS 1073  
Db 1061 SPQADMPVSTAGGTGCTIQLPEQKAIVFETGE--AGSNLGAGTLPGEVRTSTEEAT 1118  
Qy 1074 YTOETPKPSVSGISLGLPQOESAK--SATLPYIKQEEFSPRS-----QNSOPEGILL----- 1123  
Db 1119 EPQXDEVLRFTDQSL--PEDAESLSVLVSVSPDITKQEAETPRSPCSLKEQOQPHKDLMPW 1177  
Qy 1124 -----VRAQHEGVVWGTA-----CAIQEGSITRGT-----PYSK----- 1152  
Db 1178 SPEDQSLSFSEESPKETSLDISKQSLSPESLGTQLFQELNLGKEERGYPYMKAEEDSCH 1237  
Qy 1153 ISVESIPS--LRSITQGTTPALPQGTIPEALVKGIS----- 1188  
Db 1238 LAPVSIPEPHRATVSPSTDETPAGTLPGGSFSHSALSVDKHXSPGEITGPGCHFWTSDSS 1297  
Qy 1189 -----RMPIEDSSPEKREEAASKGHVYIECKSGHILSYDNIK 1226  
Db 1298 LTKSPESLSPAMEDLAVEWEGKAPGKEKEPELKSETRQKQKQILPEKVA--VVEQDLII 1355  
Qy 1227 NAREGT-----RSPRTAHEISLKRYSVESVGNTKQGMRESVPSAPLEGICRAL 1277  
Db 1356 HQKDQALDEENKPGROQDQTEPQGRDLDEKDTAAELDKGPEPKEDLDREDOG----- 1409  
Qy 1278 PRGSPHSDLKERTVLSGIMOGT---PRATESFEDGLKYPKOIK--RESPIRAFEGAI 1332  
Db 1410 QRAGPPAEKKASQORDDLOQTQATEPRDRAQERRSEEEKDKLSLELDRTTP----- 1461  
Qy 1333 TKGRPYDGITTIKEMGRSIIHPRQDILITQESR---KTPBVQOSTRPIIEGSISSQGTPI 1388  
Db 1462 -----EEKDR-----ILVQEDRAPEHSIPEPTQTDR-----APDRKGTDD 1496  
Qy 1389 KFDNNSGOSAIKHNV---KSLITGPKSLSRGMPLEIVPENIKYVERKGYEDVAKAGETVR 1445  
Db 1497 K-EQKEASEEKEQVLEQKQWALKEG-----ETLDQEARTEAO--KDETLKEDKT-Q 1545  
Qy 1446 SRHTSVVSSGSPVLRLTHLAPKAQLSPGIYDDTSA---RRTPVSYQNTMRSGSPMNRPT 1502  
Db 1546 GQKSSFVEDKTTTSKETVLQDKSAEKADSVEQQDGALEKTRALGLSESPAEGS----- 1599

Qy 1503 SDVTIPPNKSTNHERK-----STLTPTQRESIPAK--SPVPG--VDPVWVSHSPDPHHRG--- 1553  
Db 1600 -----KAREOEKKYKKEODVQGWRETSPTRCEPVGQKEPVPANEKGSPEQEVRYW 1651  
Qy 1554 -----STAGEVYWSHLPTQLDPAMPFHRALDPAAA--AYLFQRLSPPTPGVPSQQLYAM 1606  
Db 1652 RDROITLQODDAYWREL--SCDRKVMFPHELDQCGARPRYCEBERESTFDEQDEQETPL 1709  
Qy 1607 ENT--RQTLINDYITTSQOMQVNLRPDVARGLSPROPLGLP---YPATRGIIIDTINMP--- 1659  
Db 1710 QHTPRSPWTSDFKDFQEPPLPKQGLEVERWLA--ESPVLPEEEDKLTFRSFEIISPPAS 1767  
Qy 1660 -----PTILVPHPGGTSTPPMDRITYIPGTOITPPP-----RPYNASMSPG- 1701  
Db 1768 PPENTGQRPVPSAPGOESPVPDTESTAPMNEPTTPSWLAEIPPMVKDRPLPAPLSPAP 1827  
Qy 1702 -----HPTHLAAAASAERERE-----REKEKERERERI 1729  
Db 1828 APPTPAPEPHTVPVPSMGLAEYDSVAAVQEGAAELEGGPSPLGKDYRKAKGEREGEG 1887  
Qy 1730 AAASDLYLRPCSQOPGRPGSHGYVRSPSVRTQETMLQORPSVFOGTNGTSTVITPLDP 1789  
Db 1888 ACAPDSSSFSPKVPESAGESLATRTEQTEPEQREBTPPYDERSFOYADIYEQMLTLGLP 1947  
Qy 1790 TAQLRIMPPLPAGG---PSISQGLPASRYNTAADALAAALVDAASAQMDVSKTESKHEA 1846  
Db 1948 ACPTREPPLGASGDWPHLSTKEEAAGCNTSAE-----KETSSPA 1987  
Qy 1847 ARLEENLRSRAANVSEQOOLEQKTLVEKRSVQCLYTSSAFPS---CKPQPHSSVYVS- 1901  
Db 1988 S--PQNLSQDTPAFS-----YASLAGPAVPPROEPDPPGPNVEPSI 2025  
Qy 1902 -----EAGKDKGPP-----PKSRYEELRTRGKTTITTAANFIDVITIQI 1941  
Db 2026 TPVAVPPRAPISLSKDLSPPLNGSTVSCSPDRRPPSPKETGRGHWDCTNDSDL----- 2079  
Qy 1942 ASDKDARERGSO-----SDSSSSLSLSHRY-----ETPSDAI 1973  
Db 2080 --EKGAPOEPEKETRSPRPHMPMGHSSLWMPETEAYSSLSDSHLGSRVPSLDPFASAF 2137  
Qy 1974 EVISPASSPAPPOEKLOTYQPEVVKANQA-----ENDPTROYEGFLHHY 2017  
Db 2138 GFSS--LOPAPPQ-----LPSPAEPSPAPCGSLAFSGDRALALVPGTPTTRH---DEY 2186  
Qy 2018 RPOESPSPOOOLP--PSSQAEG--MGQVPRTHRLITLADHICQIITQDFARNQVSSOT 2072  
Db 2187 LEVTKAPSLDSSLQPLPSPSPGGPLLSNLR----- 2218  
Qy 2073 PQOPPTSTFQNSPALSALVSTPVRTKTSNRYSP-----ESQAQSVHHORPG 2116  
Db 2219 --PASPALSESGSSSEATTPISSVAERFPFPLEAAEQSAEGLSGSGESAHSWLDLTP 2275  
Qy 2117 SRV-----SPENLVDKSRGSRPGKSPERSHVSS 2144  
Db 2276 SPAPASLIDLAPAPAPAPAPAGPLGDLGDGTLPFCRPECTGELTKKPSFPLSPSGDHEAN 2335  
Qy 2145 EPYE--PISPPQ--VPVWHEKQDSLILLS--ORGAEPAEQORNDARSPCGSIYLSFFFTKLENT 2201  
Db 2336 GPGETSLNPPGQFVTATAKEKEEAPHAWERSWPBGAEBSRSPD-----TLLSE 2385  
Qy 2202 SPMVKSKKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSSRGHSFA--DPASN 2260  
Db 2386 QPLRPGK-----SSGG-----PPCSILSSEVEAGPOGCAATDRPH 2419  
Qy 2261 LGLEDITRKALMGSFDDKVEDHGVVMSQPMGVPGTANTSVVTSGET-----RREGRD 2313  
Db 2420 CG-----ELSPSF-----LNPLP---PPSTDSDLSTEERLAGKGRRRVGR 2459  
Qy 2314 PSHSGGVC-----KPKLISKNSRKSCKSPIGO-----GYLTE 2348  
Db 2460 PGA--TGGCPMADETPPTTSASDSGSSSDVPPPETEECPSTIAEALDSDEGDFLVD 2518  
Qy 2349 RPSVSSVHSEGDYHRQTPPGWAWEDRPSSTGTSTQTFPPNPLTMRMLSSPTPTPIAC----- 2403



```
Db 2519 KAGVSGTHPRGHPPTPLDPPRPS-----APSAVNOAAPHQONRI-WEREPAP 2426
Qy 2404 -----PPRDPVCMADPE 2558
Db 2559 GLSSEGRVERLKRGRPRGRAPRAKPAAPARRLDINGKRSPTP 2603
Qy 2404 -----PPRDPVCMADPE 2558

RESULT 8
MAPB_MOUSE
ID MAPB_MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2) [MAP1(X)] [CONTAINS: MAP1
DE LIGHT CHAIN LC1].
GN MAP1B OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "the microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT and tau.";
RL J. Cell Biol. 109:3367-3376(1989).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
CC KKEE AND KKEI/V. REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
CC BOTH IN VITRO AND IN VIVO.
CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -1- SIMILARITY: TO NEURAXIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51396; CAA35761.1;
CC PIR: S07549; QRMSP1.
CC MGD; MGI:97179; Mtap5.
CC InterPro: IPR000102;
CC Pfam: PF00414; MAP1B_neuraxin; 7.
CC PROSITE: PS00230; MAP1B_NEURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT DOMAIN 589 787
FT
FT DOMAIN 1865 2068
FT REPEAT 1865 1881
FT REPEAT 1882 1898
FT REPEAT 1899 1915
FT REPEAT 1916 1932
FT REPEAT 1933 1949
```

```
FT REPEAT 1950 1966
FT REPEAT 1967 1983
FT REPEAT 1984 2000
FT REPEAT 2001 2017
FT REPEAT 2018 2034
FT REPEAT 2035 2051
FT REPEAT 2052 2068
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 3.5%; Score 447; DB 1; Length 2464;
Best Local Similarity 18.6%; Pred. No. 1.5e-09;
Matches 490; Conservative 375; Mismatches 957; Indels 812; Gaps 116;

Qy 71 LLSFPHGSDRPPQRRRTSYEPFHGCPSPVDHDSLESRRPRLEQVSDSHFQRV---SAAVL 127
Db 174 LLSTTHPAN-----KASLTLCFPC-----EEDGWNNSNLDNRHNLQDFINIKLSASIL 220
Qy 128 PLVHPLPEGLRASADAKDPAGFGKHEAPSSPIG-----OPC-----GDD----- 168
Db 221 PEMEGLSEFTEYLSSEVPS---PFDILEPPTSGGFLKLSKPCYIPPGRGDSALFAP 277
Qy 169 -----QNASPSKLSKEELIQSDRVD-----REIAKVEQQ--- 198
Db 278 NGFNMLINGGSEKSCFWKLIRHLDRVDSILLTHIGDNDLPNGINSMLQKLALEEEERSQ 337
Qy 199 -----TLKLLKKQQ-----OLEEAAKPPPEP 219
Db 338 GSTSNSDMKNLISPDGLGVFLNVPENLKDPENIKMKRSTEEACFTLQYLNKLSMKPEP 397
Qy 220 -----EKPV-----SPPVBOKHRSIVQIIVYDENRKAEEAHKIFE 255
Db 398 LFRSVGNTEIEPVILFQKMGVCKLEMYVLNPKVSKSEKEMQYFMQOWTGNKDAE---LIL 453
Qy 256 GLGPKVELPL--YNQPSDTKYVHENIKTNQVMR-----KKLILFFKR 295
Db 454 PNGQVEDIPISYLTSSVSLIWHHPANPAEKIIRVLPNGNSTOYNILEGLEKHLKHLDFLQ 513
Qy 296 RNHARKQ-----REQKICORYDOLMEAWKKVVDRIENNPRRKAKESKT-REYEEK 344
Db 514 PLATQKDLTGTPPPVPPVQVKLQKQAD-----SRESLKATPKPVASKSVKRESKE 563
Qy 345 QFPPIRQREQEQRFQVQORGAGLSATIAKSEHSEIIDIIGLSEQNNENKQRLSVIP 404
Db 564 ETPEVTK-TSQVEKTPKVESKEKVL---VKKDKPVKTESKPSVTEKESVSKEQS----- 614
Qy 405 PMFDAAQRRVKFINMGLMEDPMKVKYKDRQFMVWTDHEKEIFDKPIQHPKPNGLIAS 464
Db 615 PVKAEVAEKQA-----TESPKVTKDKVVKKE-----IKT 644
Qy 465 YLERKSVDPDCVLYYYLTKKNENKALVRRNYGKRRGRNQOIARPSOEKVEKEEDKAEK 524
Db 645 KLEEK-----KEEKPKVEVKK-----EDKTPLKDKDKPRK 675
Qy 525 TERKEEEKDEEKEDEKSKENTKEDKIDGTAETEEREQATPRGKRTANSQRRKGR 584
Db 676 EEVKKEIKKEIKERKELKKEVKKETPLDKAKKEVKEEKEKVEKKEKPKKIKKISK 735
Qy 585 ITRSWTNEAAAAAASAAAAATEPPPLPPPEPDPISLTPVETSRWTEE-EMEVAKKGLVGH 643
Db 736 DIKSTPQSDTKKPSAL-----KPKVAKKEESTKKEPLAAGKLKDKGKVKVKK----- 785
Qy 644 GRNWAATAKMWGTKSEAOCKNFYFNKRRHNLNLLQOHKQKTSRKPREEDVSOCSVA 703
Db 786 GKTEAAATAGTAATTA-----VVAAGIAASGPVKELEAER 824
Qy 704 STVSAQDEDEIEASNEENPDSEV-EAVKPSDESPENATSRGNTPE-----AVELEPTE 758
Db 825 SLNSS--PEDLTDFEELKAEIDVAKDIKPOLIEDEEKLKTKETQPCAVVIQKETEVS 882
Qy 759 TAPSTSPSLAVPSTK-----PADESVEYQVNDSI-----SAET----- 792
Db 883 KGSAESPDGEGITTEGEGEGCEQTPPELEPVEKQGVDDIDKFEDEGAGFESSETGTYEEK 942
```























DR Pfam; PF00531; death; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 20.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50017; DEATH\_DOMAIN; 1.  
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;  
KW Phosphorylation; Multigene family;  
FT REPEAT 63 92  
FT REPEAT ANK 1.  
FT REPEAT 96 125  
FT REPEAT 129 158  
FT REPEAT 162 191  
FT REPEAT 193 220  
FT REPEAT 232 261  
FT REPEAT 265 294  
FT REPEAT 298 327  
FT REPEAT 331 360  
FT REPEAT 364 393  
FT REPEAT 397 426  
FT REPEAT 430 459  
FT REPEAT 463 492  
FT REPEAT 496 525  
FT REPEAT 529 558  
FT REPEAT 562 591  
FT REPEAT 595 624  
FT REPEAT 628 657  
FT REPEAT 661 690  
FT REPEAT 694 723  
FT REPEAT 727 756  
FT REPEAT 760 789  
FT REPEAT 793 822  
FT REPEAT 825 854  
FT REPEAT 857 886  
FT REPEAT 889 918  
FT REPEAT 921 950  
FT REPEAT 953 982  
FT REPEAT 985 1014  
FT REPEAT 1017 1046  
FT REPEAT 1049 1078  
FT REPEAT 1081 1110  
FT REPEAT 1113 1142  
FT REPEAT 1145 1174  
FT REPEAT 1177 1206  
FT REPEAT 1209 1238  
FT REPEAT 1241 1270  
FT REPEAT 1273 1302  
FT REPEAT 1305 1334  
FT REPEAT 1337 1366  
FT REPEAT 1369 1398  
FT REPEAT 1401 1430  
FT REPEAT 1433 1462  
FT REPEAT 1465 1494  
FT REPEAT 1497 1526  
FT REPEAT 1529 1558  
FT REPEAT 1561 1590  
FT REPEAT 1593 1622  
FT REPEAT 1625 1654  
FT REPEAT 1657 1686  
FT REPEAT 1689 1718  
FT REPEAT 1721 1750  
FT REPEAT 1753 1782  
FT REPEAT 1785 1814  
FT REPEAT 1817 1846  
FT REPEAT 1849 1878  
FT REPEAT 1881 1910  
FT REPEAT 1913 1942  
FT REPEAT 1945 1974  
FT REPEAT 1977 2006  
FT REPEAT 2009 2038  
FT REPEAT 2041 2070  
FT REPEAT 2073 2102  
FT REPEAT 2105 2134  
FT REPEAT 2137 2166  
FT REPEAT 2169 2198  
FT REPEAT 2201 2230  
FT REPEAT 2233 2262  
FT REPEAT 2265 2294  
FT REPEAT 2297 2326  
FT REPEAT 2329 2358  
FT REPEAT 2361 2390  
FT REPEAT 2393 2422  
FT REPEAT 2425 2454  
FT REPEAT 2457 2486  
FT REPEAT 2489 2518  
FT REPEAT 2521 2550  
FT REPEAT 2553 2582  
FT REPEAT 2585 2614  
FT REPEAT 2617 2646  
FT REPEAT 2649 2678  
FT REPEAT 2681 2710  
FT REPEAT 2713 2742  
FT REPEAT 2745 2774  
FT REPEAT 2777 2806  
FT REPEAT 2809 2838  
FT REPEAT 2841 2870  
FT REPEAT 2873 2902  
FT REPEAT 2905 2934  
FT REPEAT 2937 2966  
FT REPEAT 2969 2998  
FT REPEAT 3001 3030  
FT REPEAT 3033 3062  
FT REPEAT 3065 3094  
FT REPEAT 3097 3126  
FT REPEAT 3129 3158  
FT REPEAT 3161 3190  
FT REPEAT 3193 3222  
FT REPEAT 3225 3254  
FT REPEAT 3257 3286  
FT REPEAT 3289 3318  
FT REPEAT 3321 3350  
FT REPEAT 3353 3382  
FT REPEAT 3385 3414  
FT REPEAT 3417 3446  
FT REPEAT 3449 3478  
FT REPEAT 3481 3510  
FT REPEAT 3513 3542  
FT REPEAT 3545 3574  
FT REPEAT 3577 3606  
FT REPEAT 3609 3638  
FT REPEAT 3641 3670  
FT REPEAT 3673 3702  
FT REPEAT 3705 3734  
FT REPEAT 3737 3766  
FT REPEAT 3769 3798  
FT REPEAT 3801 3830  
FT REPEAT 3833 3862  
FT REPEAT 3865 3894  
FT REPEAT 3897 3926  
FT REPEAT 3929 3958  
FT REPEAT 3961 3990  
FT REPEAT 3993 4022  
FT REPEAT 4025 4054  
FT REPEAT 4057 4086  
FT REPEAT 4089 4118  
FT REPEAT 4121 4150  
FT REPEAT 4153 4182  
FT REPEAT 4185 4214  
FT REPEAT 4217 4246  
FT REPEAT 4249 4278  
FT REPEAT 4281 4310  
FT REPEAT 4313 4342  
FT REPEAT 4345 4374  
FT REPEAT 4377 4406  
FT REPEAT 4409 4438  
FT REPEAT 4441 4470  
FT REPEAT 4473 4502  
FT REPEAT 4505 4534  
FT REPEAT 4537 4566  
FT REPEAT 4569 4598  
FT REPEAT 4601 4630  
FT REPEAT 4633 4662  
FT REPEAT 4665 4694  
FT REPEAT 4697 4726  
FT REPEAT 4729 4758  
FT REPEAT 4761 4790  
FT REPEAT 4793 4822  
FT REPEAT 4825 4854  
FT REPEAT 4857 4886  
FT REPEAT 4889 4918  
FT REPEAT 4921 4950  
FT REPEAT 4953 4982  
FT REPEAT 4985 5014  
FT REPEAT 5017 5046  
FT REPEAT 5049 5078  
FT REPEAT 5081 5110  
FT REPEAT 5113 5142  
FT REPEAT 5145 5174  
FT REPEAT 5177 5206  
FT REPEAT 5209 5238  
FT REPEAT 5241 5270  
FT REPEAT 5273 5302  
FT REPEAT 5305 5334  
FT REPEAT 5337 5366  
FT REPEAT 5369 5398  
FT REPEAT 5401 5430  
FT REPEAT 5433 5462  
FT REPEAT 5465 5494  
FT REPEAT 5497 5526  
FT REPEAT 5529 5558  
FT REPEAT 5561 5590  
FT REPEAT 5593 5622  
FT REPEAT 5625 5654  
FT REPEAT 5657 5686  
FT REPEAT 5689 5718  
FT REPEAT 5721 5750  
FT REPEAT 5753 5782  
FT REPEAT 5785 5814  
FT REPEAT 5817 5846  
FT REPEAT 5849 5878  
FT REPEAT 5881 5910  
FT REPEAT 5913 5942  
FT REPEAT 5945 5974  
FT REPEAT 5977 6006  
FT REPEAT 6009 6038  
FT REPEAT 6041 6070  
FT REPEAT 6073 6102  
FT REPEAT 6105 6134  
FT REPEAT 6137 6166  
FT REPEAT 6169 6198  
FT REPEAT 6201 6230  
FT REPEAT 6233 6262  
FT REPEAT 6265 6294  
FT REPEAT 6297 6326  
FT REPEAT 6329 6358  
FT REPEAT 6361 6390  
FT REPEAT 6393 6422  
FT REPEAT 6425 6454  
FT REPEAT 6457 6486  
FT REPEAT 6489 6518  
FT REPEAT 6521 6550  
FT REPEAT 6553 6582  
FT REPEAT 6585 6614  
FT REPEAT 6617 6646  
FT REPEAT 6649 6678  
FT REPEAT 6681 6710  
FT REPEAT 6713 6742  
FT REPEAT 6745 6774  
FT REPEAT 6777 6806  
FT REPEAT 6809 6838  
FT REPEAT 6841 6870  
FT REPEAT 6873 6902  
FT REPEAT 6905 6934  
FT REPEAT 6937 6966  
FT REPEAT 6969 6998  
FT REPEAT 7001 7030  
FT REPEAT 7033 7062  
FT REPEAT 7065 7094  
FT REPEAT 7097 7126  
FT REPEAT 7129 7158  
FT REPEAT 7161 7190  
FT REPEAT 7193 7222  
FT REPEAT 7225 7254  
FT REPEAT 7257 7286  
FT REPEAT 7289 7318  
FT REPEAT 7321 7350  
FT REPEAT 7353 7382  
FT REPEAT 7385 7414  
FT REPEAT 7417 7446  
FT REPEAT 7449 7478  
FT REPEAT 7481 7510  
FT REPEAT 7513 7542  
FT REPEAT 7545 7574  
FT REPEAT 7577 7606  
FT REPEAT 7609 7638  
FT REPEAT 7641 7670  
FT REPEAT 7673 7702  
FT REPEAT 7705 7734  
FT REPEAT 7737 7766  
FT REPEAT 7769 7798  
FT REPEAT 7801 7830  
FT REPEAT 7833 7862  
FT REPEAT 7865 7894  
FT REPEAT 7897 7926  
FT REPEAT 7929 7958  
FT REPEAT 7961 7990  
FT REPEAT 7993 8022  
FT REPEAT 8025 8054  
FT REPEAT 8057 8086  
FT REPEAT 8089 8118  
FT REPEAT 8121 8150  
FT REPEAT 8153 8182  
FT REPEAT 8185 8214  
FT REPEAT 8217 8246  
FT REPEAT 8249 8278  
FT REPEAT 8281 8310  
FT REPEAT 8313 8342  
FT REPEAT 8345 8374  
FT REPEAT 8377 8406  
FT REPEAT 8409 8438  
FT REPEAT 8441 8470  
FT REPEAT 8473 8502  
FT REPEAT 8505 8534  
FT REPEAT 8537 8566  
FT REPEAT 8569 8598  
FT REPEAT 8601 8630  
FT REPEAT 8633 8662  
FT REPEAT 8665 8694  
FT REPEAT 8697 8726  
FT REPEAT 8729 8758  
FT REPEAT 8761 8790  
FT REPEAT 8793 8822  
FT REPEAT 8825 8854  
FT REPEAT 8857 8886  
FT REPEAT 8889 8918  
FT REPEAT 8921 8950  
FT REPEAT 8953 8982  
FT REPEAT 8985 9014  
FT REPEAT 9017 9046  
FT REPEAT 9049 9078  
FT REPEAT 9081 9110  
FT REPEAT 9113 9142  
FT REPEAT 9145 9174  
FT REPEAT 9177 9206  
FT REPEAT 9209 9238  
FT REPEAT 9241 9270  
FT REPEAT 9273 9302  
FT REPEAT 9305 9334  
FT REPEAT 9337 9366  
FT REPEAT 9369 9398  
FT REPEAT 9401 9430  
FT REPEAT 9433 9462  
FT REPEAT 9465 9494  
FT REPEAT 9497 9526  
FT REPEAT 9529 9558  
FT REPEAT 9561 9590  
FT REPEAT 9593 9622  
FT REPEAT 9625 9654  
FT REPEAT 9657 9686  
FT REPEAT 9689 9718  
FT REPEAT 9721 9750  
FT REPEAT 9753 9782  
FT REPEAT 9785 9814  
FT REPEAT 9817 9846  
FT REPEAT 9849 9878  
FT REPEAT 9881 9910  
FT REPEAT 9913 9942  
FT REPEAT 9945 9974  
FT REPEAT 9977 10006  
FT REPEAT 10009 10038  
FT REPEAT 10041 10070  
FT REPEAT 10073 10102  
FT REPEAT 10105 10134  
FT REPEAT 10137 10166  
FT REPEAT 10169 10198  
FT REPEAT 10201 10230  
FT REPEAT 10233 10262  
FT REPEAT 10265 10294  
FT REPEAT 10297 10326  
FT REPEAT 10329 10358  
FT REPEAT 10361 10390  
FT REPEAT 10393 10422  
FT REPEAT 10425 10454  
FT REPEAT 10457 10486  
FT REPEAT 10489 10518  
FT REPEAT 10521 10550  
FT REPEAT 10553 10582  
FT REPEAT 10585 10614  
FT REPEAT 10617 10646  
FT REPEAT 10649 10678  
FT REPEAT 10681 10710  
FT REPEAT 10713 10742  
FT REPEAT 10745 10774  
FT REPEAT 10777 10806  
FT REPEAT 10809 10838  
FT REPEAT 10841 10870  
FT REPEAT 10873 10902  
FT REPEAT 10905 10934  
FT REPEAT 10937 10966  
FT REPEAT 10969 10998  
FT REPEAT 11001 11030  
FT REPEAT 11033 11062  
FT REPEAT 11065 11094  
FT REPEAT 11097 11126  
FT REPEAT 11129 11158  
FT REPEAT 11161 11190  
FT REPEAT 11193 11222  
FT REPEAT 11225 11254  
FT REPEAT 11257 11286  
FT REPEAT 11289 11318  
FT REPEAT 11321 11350  
FT REPEAT 11353 11382  
FT REPEAT 11385 11414  
FT REPEAT 11417 11446  
FT REPEAT 11449 11478  
FT REPEAT 11481 11510  
FT REPEAT 11513 11542  
FT REPEAT 11545 11574  
FT REPEAT 11577 11606  
FT REPEAT 11609 11638  
FT REPEAT 11641 11670  
FT REPEAT 11673 11702  
FT REPEAT 11705 11734  
FT REPEAT 11737 11766  
FT REPEAT 11769 11798  
FT REPEAT 11801 11830  
FT REPEAT 11833 11862  
FT REPEAT 11865 11894  
FT REPEAT 11897 11926  
FT REPEAT 11929 11958  
FT REPEAT 11961 11990  
FT REPEAT 11993 12022  
FT REPEAT 12025 12054  
FT REPEAT 12057 12086  
FT REPEAT 12089 12118  
FT REPEAT 12121 12150  
FT REPEAT 12153 12182  
FT REPEAT 12185 12214  
FT REPEAT 12217 12246  
FT REPEAT 12249 12278  
FT REPEAT 12281 12310  
FT REPEAT 12313 12342  
FT REPEAT 12345 12374  
FT REPEAT 12377 12406  
FT REPEAT 12409 12438  
FT REPEAT 12441 12470  
FT REPEAT 12473 12502  
FT REPEAT 12505 12534  
FT REPEAT 12537 12566  
FT REPEAT 12569 12598  
FT REPEAT 12601 12630  
FT REPEAT 12633 12662  
FT REPEAT 12665 12694  
FT REPEAT 12697 12726  
FT REPEAT 12729 12758  
FT REPEAT 12761 12790  
FT REPEAT 12793 12822  
FT REPEAT 12825 12854  
FT REPEAT 12857 12886  
FT REPEAT 12889 12918  
FT REPEAT 12921 12950  
FT REPEAT 12953 12982  
FT REPEAT 12985 13014  
FT REPEAT 13017 13046  
FT REPEAT 13049 13078  
FT REPEAT 13081 13110  
FT REPEAT 13113 13142  
FT REPEAT 13145 13174  
FT REPEAT 13177 13206  
FT REPEAT 13209 13238  
FT REPEAT 13241 13270  
FT REPEAT 13273 13302  
FT REPEAT 13305 13334  
FT REPEAT 13337 13366  
FT REPEAT 13369 13398  
FT REPEAT 13401 13430  
FT REPEAT 13433 13462  
FT REPEAT 13465 13494  
FT REPEAT 13497 13526  
FT REPEAT 13529 13558  
FT REPEAT 13561 13590  
FT REPEAT 13593 13622  
FT REPEAT 13625 13654  
FT REPEAT 13657 13686  
FT REPEAT 13689 13718  
FT REPEAT 13721 13750  
FT REPEAT 13753 13782  
FT REPEAT 13785 13814  
FT REPEAT 13817 13846  
FT REPEAT 13849 13878  
FT REPEAT 13881 13910  
FT REPEAT 13913 13942  
FT REPEAT 13945 13974  
FT REPEAT 13977 14006  
FT REPEAT 14009 14038  
FT REPEAT 14041 14070  
FT REPEAT 14073 14102  
FT REPEAT 14105 14134  
FT REPEAT 14137 14166  
FT REPEAT 14169 14198  
FT REPEAT 14201 14230  
FT REPEAT 14233 14262  
FT REPEAT 14265 14294  
FT REPEAT 14297 14326  
FT REPEAT 14329 14358  
FT REPEAT 14361 14390  
FT REPEAT 14393 14422  
FT REPEAT 14425 14454  
FT REPEAT 14457 14486  
FT REPEAT 14489 14518  
FT REPEAT 14521 14550  
FT REPEAT 14553 14582  
FT REPEAT 14585 14614  
FT REPEAT 14617 14646  
FT REPEAT 14649 14678  
FT REPEAT 14681 14710  
FT REPEAT 14713 14742  
FT REPEAT 14745 14774  
FT REPEAT 14777 14806  
FT REPEAT 14809 14838  
FT REPEAT 14841 14870  
FT REPEAT 14873 14902  
FT REPEAT 14905 14934  
FT REPEAT 14937 14966  
FT REPEAT 14969 14998  
FT REPEAT 15001 15030  
FT REPEAT 15033 15062  
FT REPEAT 15065 15094  
FT REPEAT 15097 15126  
FT REPEAT 15129 15158  
FT REPEAT 15161 15190  
FT REPEAT 15193 15222  
FT REPEAT 15225 15254  
FT REPEAT 15257 15286  
FT REPEAT 15289 15318  
FT REPEAT 15321 15350  
FT REPEAT 15353 15382  
FT REPEAT 15385 15414  
FT REPEAT 15417 15446  
FT REPEAT 15449 15478  
FT REPEAT 15481 15510  
FT REPEAT 15513 15542  
FT REPEAT 15545 15574  
FT REPEAT 15577 15606  
FT REPEAT 15609 15638  
FT REPEAT 15641 15670  
FT REPEAT 15673 15702  
FT REPEAT 15705 15734  
FT REPEAT 15737 15766  
FT REPEAT 15769 15798  
FT REPEAT 15801 15830  
FT REPEAT 15833 15862  
FT REPEAT 15865 15894  
FT REPEAT 15897 15926  
FT REPEAT 15929 15958  
FT REPEAT 15961 15990  
FT REPEAT 15993 16022  
FT REPEAT 16025 16054  
FT REPEAT 16057 16086  
FT REPEAT 16089 16118  
FT REPEAT 16121 16150  
FT REPEAT 16153 16182  
FT REPEAT 16185 16214  
FT REPEAT 16217 16246  
FT REPEAT 16249 16278  
FT REPEAT 16281 16310  
FT REPEAT 16313 16342  
FT REPEAT 16345 16374  
FT REPEAT 16377 16406  
FT REPEAT 16409 16438  
FT REPEAT 16441 16470  
FT REPEAT 16473 16502  
FT REPEAT 16505 16534  
FT REPEAT 16537 16566  
FT REPEAT 16569 16598  
FT REPEAT 16601 16630  
FT REPEAT 16633 16662  
FT REPEAT 16665 16694  
FT REPEAT 16697 16726  
FT REPEAT 16729 16758  
FT REPEAT 16761 16790  
FT REPEAT 16793 16822  
FT REPEAT 16825 16854  
FT REPEAT 16857 16886  
FT REPEAT 16889 16918  
FT REPEAT 16921 16950  
FT REPEAT 16953 16982  
FT REPEAT 16985 17014  
FT REPEAT 17017 17046  
FT REPEAT 17049 17078  
FT REPEAT 17081 17110  
FT REPEAT 17113 17142  
FT REPEAT 17145 17174  
FT REPEAT 17177 17206  
FT REPEAT 17209 17238  
FT REPEAT 17241 17270  
FT REPEAT 17273 17302  
FT REPEAT 17305 17334  
FT REPEAT 17337 17366  
FT REPEAT 17369 17398  
FT REPEAT 17401 17430  
FT REPEAT 17433 17462  
FT REPEAT 17465 17494  
FT REPEAT 17497 17526  
FT REPEAT 17529 17558  
FT REPEAT 17561 17590  
FT REPEAT 17593 17622  
FT REPEAT 17625 17654  
FT REPEAT 17657 17686  
FT REPEAT 17689 17718  
FT REPEAT 17721 17750  
FT REPEAT 17753 17782  
FT REPEAT 17785 17814  
FT REPEAT 17817 17846  
FT REPEAT 17849 17878  
FT REPEAT 17881 17910  
FT REPEAT 17913 17942  
FT REPEAT 17945 17974  
FT REPEAT 17977 18006  
FT REPEAT 18009 18038  
FT REPEAT 18041 18070  
FT REPEAT 18073 18102  
FT REPEAT 18105 18134  
FT REPEAT 18137 18166  
FT REPEAT 18169 18198  
FT REPEAT 18201 18230  
FT REPEAT 18233 18262  
FT REPEAT 18265 18294  
FT REPEAT 18297 18326  
FT REPEAT 18329 18358  
FT REPEAT 18361 18390  
FT REPEAT 18393 18422  
FT REPEAT 18425 18454  
FT REPEAT 18457 18486  
FT REPEAT 18489 18518  
FT REPEAT 18521 18550  
FT REPEAT 18553 18582  
FT REPEAT 18585 18614  
FT REPEAT 18617 18646  
FT REPEAT 18649 18678  
FT REPEAT 18681 18710  
FT REPEAT 18713 18742  
FT REPEAT 18745 18774  
FT REPEAT 18777 18806  
FT REPEAT 18809 18838  
FT REPEAT 18841 18870  
FT REPEAT 18873 18902  
FT REPEAT 18905 18934  
FT REPEAT 18937 18966  
FT REPEAT 18969 18998  
FT REPEAT 19001 19030  
FT REPEAT 19033 19062  
FT REPEAT 19065 19094  
FT REPEAT 19097 19126  
FT REPEAT 19129 19158  
FT REPEAT 19161 19190  
FT REPEAT 19193 19222  
FT REPEAT 19225 19254  
FT REPEAT 19257 19286  
FT REPEAT 19289 19318  
FT REPEAT 19321 19350  
FT REPEAT 19353 19382  
FT REPEAT 19385 19414  
FT REPEAT 19417 19446  
FT REPEAT 19449 19478  
FT REPEAT 19481 19510  
FT REPEAT 19513 19542  
FT REPEAT 19545 19574  
FT REPEAT 19577 19606  
FT REPEAT 19609 19638  
FT REPEAT 19641 19670  
FT REPEAT 19673 19702  
FT REPEAT 19705 19734  
FT REPEAT 19737 19766  
FT REPEAT 19769 19798  
FT REPEAT 19801 19830  
FT REPEAT 19833 19862  
FT REPEAT 19865 19894  
FT REPEAT 19897 19926  
FT REPEAT 19929 19958  
FT REPEAT 19961 19990  
FT REPEAT 19993 20022  
FT REPEAT 20025 20054  
FT REPEAT 20057 20086  
FT REPEAT 20089 20118  
FT REPEAT 20121 20150  
FT REPEAT 20153 20182  
FT REPEAT 20185 20214  
FT REPEAT 20217 20246  
FT REPEAT 20249 20278  
FT REPEAT 20281 20310  
FT REPEAT 20313 20342  
FT REPEAT 20345 20374  
FT REPEAT 20377 20406  
FT REPEAT 20409 20438  
FT REPEAT 20441 20470  
FT REPEAT 20473 20502  
FT REPEAT 20505 20534  
FT REPEAT 20537 20566  
FT REPEAT 20569 20598  
FT REPEAT 20601 20630  
FT REPEAT 20633 20662  
FT REPEAT 20665 20694  
FT REPEAT 20697 20726  
FT REPEAT 20729 20758  
FT REPEAT 20761 20790  
FT REPEAT 20793 20822  
FT REPEAT 20825 20854  
FT REPEAT 20857 20886  
FT REPEAT 20889 20918  
FT REPEAT 20921 20950  
FT REPEAT 20953 20982  
FT REPEAT 20985 21014  
FT REPEAT 21017 21046  
FT REPEAT 21049 21078  
FT REPEAT 21081 21110  
FT REPEAT 21113 21142  
FT REPEAT 21145 21174  
FT REPEAT 21177 21206  
FT REPEAT 21209 21238  
FT REPEAT 21241 21270  
FT REPEAT 21273 21302  
FT REPEAT 21305 21334  
FT REPEAT 21337 21366  
FT REPEAT 21369 21398  
FT REPEAT 21401 21430  
FT REPEAT 21433 21462  
FT REPEAT 21465 21494  
FT REPEAT 21497 21526  
FT REPEAT 21529 21558  
FT REPEAT 21561 21590  
FT REPEAT 21593 21622  
FT REPEAT 21625 21654  
FT REPEAT 21657 21686  
FT REPEAT 21689 21718  
FT REPEAT 21721 21750  
FT REPEAT 21753 21782  
FT REPEAT 21785 21814  
FT REPEAT 21817 21846  
FT REPEAT 21849 21878  
FT REPEAT 21881 21910  
FT REPEAT 21913 21942  
FT REPEAT 21945 21974  
FT REPEAT 21977 22006  
FT REPEAT 22009 22038  
FT REPEAT 22041 22070  
FT REPEAT 22073 22102  
FT REPEAT 22105 22134  
FT REPEAT 22137 22166  
FT REPEAT 22169 22198  
FT REPEAT 22201 22230  
FT REPEAT 22233 22262  
FT REPEAT 22265 22294  
FT REPEAT 22297 22326  
FT REPEAT 22329 22358  
FT REPEAT 22361 22390  
FT REPEAT 22393 22422  
FT REPEAT 22425 22454  
FT REPEAT 22457 22486  
FT REPEAT 22489 22518  
FT REPEAT 22521 22550  
FT REPEAT 22553 22582  
FT REPEAT 22585 22614  
FT REPEAT 22617 22646  
FT REPEAT 22649 22678  
FT REPEAT 22681 22710  
FT REPEAT 22713 22742  
FT REPEAT 22745 22774  
FT REPEAT 22777 22806  
FT REPEAT 22809 22838  
FT REPEAT 22841 22870  
FT REPEAT 22873 22902  
FT REPEAT 22905 22934  
FT REPEAT 22937 22966  
FT REPEAT 22969 22998  
FT REPEAT 23001 23030  
FT REPEAT 23033 23062  
FT REPEAT 23065 23094  
FT REPEAT 23097 23126  
FT REPEAT 23129 23158  
FT REPEAT 23161 23190  
FT REPEAT 23193 23222  
FT REPEAT 23225 23254  
FT REPEAT 23257 23286  
FT REPEAT 23289 23318  
FT REPEAT 23321 23350  
FT REPEAT 23353 23382  
FT REPEAT 23385 23414  
FT REPEAT 23417 23446  
FT REPEAT 23449 23478  
FT REPEAT 23481 23510  
FT REPEAT 23513 23542  
FT REPEAT 23545 23574  
FT REPEAT 23577 23606  
FT REPEAT 23609 23638  
FT REPEAT 23641 23670  
FT REPEAT 23673 23702  
FT REPEAT 23705 23734  
FT REPEAT 23737 23766  
FT REPEAT 23769 23798  
FT REPEAT 23801 23830  
FT REPEAT 23833 23862  
FT REPEAT 23865 23894  
FT REPEAT 23897 23926  
FT REPEAT 23929 23958  
FT REPEAT 23961 23990  
FT REPEAT 23993 24022  
FT REPEAT 24025 24054  
FT REPEAT 24057 24086  
FT REPEAT 24089 24118  
FT REPEAT 24121



Db 2833 -SATSIQKTEVTKTDTETENLPKPCQSDSSITTTQDRFSMD-- 2874  
Qy 1178 PTEALVKGSI SMPLEDSPEKGRGAAKSHVYEGKSHLS-YDNI-KNAREGTRSP 1235  
Db 2875 -VPVSD--LAENDEYDQ---ITSPYENVPQSQFSSBES 2909  
Qy 1236 RTAHEISLKRYSVEGNKQGMNRESVPAPLEGLICRALPRGSPHS-----DLK 1287  
Db 2910 KTQTDANHTTFSHSE-----VYSVTITSPVEDVVVASSSGTVLSKESNPEGQDIK 2961  
Qy 1288 ERTVLSSIMOGTFRATTESTEDGLKPKQIKRESPPIRAFEGAITK-GKPYDGIITIKE 1346  
Db 2962 MESOLETLWQMDSVSSSEPTMSATTVVGE-----QISKVIITKTDVDSWSEIRE 3017  
Qy 1347 MGRSITHEIPROD-----ILTQESRKTPEVQSTRIEES-ISQGTPIKFD----- 1391  
Db 3018 DDEAFEARKEEOKIFGLVDRQSGTTPPTPARTTEEGTPTSEONPFLFOEGKLF 3077  
Qy 1392 -NNSGSAIKHNKSLITGPSKLSRGMPPLEIIVPEN 1426  
Db 3078 MTRSGAIDMKRSVADESFFHQIGQESREETLSEDKVKGATGADPL-----PLETSAES 3132  
Qy 1427 IKVVERGKYEDVKAQETVRSRHTSVSGSPVLSRLTLEAP--KAQLSP--GIYDDTSAR 1482  
Db 3133 LALSE-----SKETVDDDEADLL-----PDSVSEEEVEIPASDAOLNSQMG---SAST 3177  
Qy 1483 RTPVSYQNTMSRSGPMNRTSDVTIPPNKSTNHERKSTLTPTQRESIPAKSPVPGVDPVV 1542  
Db 3178 ETPT--KEAVSVG-----TKD-----LPTVQGDIP---PLSGVQKI-- 3209  
Qy 1543 SHSPFDPHHRGSTAGEVYVSHLPTQLDPAMPFHRALDPAAAYLFRQLSPPTPGYPSQY 1602  
Db 3210 -SCPDSSEPAV-----QVQ 3222  
Qy 1603 LYAMENTROTILNDYITISQOMQVNLDPDVARGLSPREQ-----PL-CLPYPARGI 1652  
Db 3223 L-----DFSLTRSVSDRGDSDSPSSPEEQKSVIEIPTAPMENVPFTESKSK 3270  
Qy 1653 IDLTNMPPTILVPHGGTSTPPMDRITVIGTQITFPPRPNYSASMSGHPHTLAAAASA 1712  
Db 3271 IPVRTMP-----TSTPA-----PPSAEYESVSEDF-----LSSV 3300  
Qy 1713 EREREREKERERERIAAASDLYLRGSDQGRPGSHGVYRSPSPSVRTQETMLOQRP 1772  
Db 3301 DEENKADEAK-----PKSKLP-----VKVPLQVREQ----- 3327  
Qy 1773 SVFOGTNCTSVITPLDPTAQLRIMPLPAGGSPISOGULPASRYNTAADAALVDAASAP 1832  
Db 3328 -LSLDTSVQKTVP-----QGDMAASIAPDNRSKSESD--ASSLDSKTACP 3371  
Qy 1833 QMDVSKTESKHEAARLEENLRSAVSEOOOLEQKTLVEKRSVOCLYTSSAFPSGKP 1892  
Db 3372 VKTRSYP-ETTESRERAELESEECATPKILTSLPKVSRSTSSCRGGTSPKES 3430  
Qy 1893 QPHSSVYSEAGDKGPPPKSKRYEEELTRGKTTITTAANFIDVITRQIASDKDARERS 1952  
Db 3431 KEHFPDLYRNS-----IEFFEETSD-----ASKLVDRL-----TQSEREQEI 3468  
Qy 1953 QSSDSSSLSHRVTSDATEIVSPASSPAPQEKIQTQYQVVKANQANDTRQYEG 2012  
Db 3469 VSDDESS-----ALEVSVLENPPVTEHSVPEDFTDTPWDESIETLIERIPENG 3522  
Qy 2013 PLHYRPOESPSPOOQLPPSSQAEGMGQVPRTHRLTLADHICQITQ-----DEARNQ 2067  
Db 3523 HDHAEDPODE-----QERI-----EERLAVIADHLGFSWTELARELDFTEEQ 3564  
Qy 2068 VSSQTPOQPPTSTFQNSPSALVSTPVRTKTSNRYSPESQ-AQSVHHQPGRSVRSPENLVD 2126  
Db 3565 IHQ-----IRENPSNLQDQSQYLLKWLWRDGRKHAFTNLV- 3601  
Qy 2127 KSGSRPCKSPERSHVSEPEYTPSPQVVPVHEKQDLSLLLSQRGAPAPBQRNDARSPG 2186  
Db 3602 -----ECLTKIN--RMDIVH-----LMEITNTEPLQER----- 3626

Qy 2187 SISYLPSPFTKLENTSPMVKSKQEIFR-KLNSSGGGSDMAAQPQGTETFNLPATVTS 2245  
Db 3627 -ISHSYAEIEOTITLDHSEGFVLOEEELCTAQHKQKQAVSKESKTCDDHPPIVSEE 3682  
Qy 2246 SVSRGHSFAD-----PASNLGL-----EDIIRKALMGSGFDDKVEDHGVVMSQPMGV- 2292  
Db 3683 DISVGYSTFQDGVPKTEGSDSSSTALFPQTHKEQVQODFSGKMODLPESSLEYQOEYFT 3742  
Qy 2293 VPGTANTSVVTSGETRRREGDPSPHSGGVCKPKLISNSRKSRSKSPGQGYLGTTERPSS 2352  
Db 3743 TPGT-----ETSETQAMIVPS-----SP-----SKTPEE 3767  
Qy 2353 VSSVHSEGDYHROTQPGMAWEDRPSSTGST---OPPNPLTMRMLSSSTPTPIACAPS 2409  
Db 3768 VSTPAEEKLYLQTP-----TSSRGSGPIIOPEPSEHRESSPRKTSLVIVESADN 3821  
Qy 2410 Q 2410  
Db 3822 Q 3822

## RESULT 12

ID APC\_RAT STANDARD; PRT; 2842 AA.  
AC P70478;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).  
GN APC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FSCHER 344/N; TISSUE=Brain;  
RX MEDLINE=96116966; Pubmed=8563176;  
RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,  
RA Sugimura T., Nagao M.;  
RT "CDNA cloning of the rat APC gene and assignment to chromosome 18.";  
RL Mamm. Genome 6:746-748(1995).  
RN [2]  
RP MUTAGENESIS.  
RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;  
RX MEDLINE=95148647; Pubmed=7846077;  
RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,  
RA Weisburger J.H., Sugimura T., Nagao M.;  
RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon  
tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).  
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-  
CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY  
SIMILARITY).  
CC -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATENINS (BY  
SIMILARITY).  
CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL; D38629; BAA07609.1;  
DR HSSP; Q02248; 2ECT.  
DR InterPro; IPR000225;  
DR Pfam; PF00514; Armadillo\_seg; 4.











Qy 309 QRYDOLMEAWKVDRIENPRKAK-----ESKREYVEKOPPEIRKOREQOERFORVGO 364  
Db 825 SASPLRRQCIENGCVAKTPTANYKTSLTKTSD-----TETESKVTSTVNR 874  
Qy 365 RGAGLSATIAKSEHSEIBIDLSQENNE-----KOMROLSVPPMFDARVRVFI 418  
Db 875 SG---RSTEFNRNIQKLP--VESKSEETWTEIVECILKRGOKATLL-----QORR----- 918  
Qy 419 NNWGLM---EDPMKYKQDQFQWVWTDHEKELFKDKFTQHPKNPGL-----TASYLERKSV 471  
Db 919 --EGEMKEIERPFETYE-----NIELKENDE--KMKAMKRSRTWQKCAPMSDLTDLKSL 970  
Qy 472 PDCVLYLYLTKNENYKALVRNRYGKRGRNQIARPSQEEKVEKEEDKAEK----- 524  
Db 971 PDTLM-----KOTARQNLQIOTODHAKAPKSEKGTITKPCQSLQ 1011  
Qy 525 -----TEKKEEKKOBEKDEKED-----SKENTREKDKIDTAET---E 562  
Db 1012 PEPINTPTHKQOLKASLGKVGVEELLAVGKFTRTSGETTHHREPAGDCKSINTPKES 1071  
Qy 563 EREQATPRKRTANSQGR-----KGRITRSTMTNEAAAAASAAA 601  
Db 1072 PKQILDPAARVTGMKKWPRTKPEEAQSLDLAGFKELFQTPCSEESMTDEKTTKIACKS 1131  
Qy 602 AATEPPPPPLPPPEPISTEPVETSRW-----TEEEMEVAKKGLVHEGRNWAALAK 652  
Db 1132 -----PPRESVDT--PTSTQWPKRSRLKADVVEEFALURLKLTPSAGK--AMLTP 1177  
Qy 653 MVGKSEAOCNFYNYKRRNLDNLQOHKOTSRRKPRE-----RVSQCESVASTVSA 708  
Db 1178 KPAGGDEKIDIAFMTPTVQKLDLAGTLPKGRQL--QTPKEKAQALEDLAGFKELFQTPG- 1235  
Qy 709 QEDIEDIASNEE-----ENPEDSEVAVKPSEDSPENATSRGTEPAVELEPTEETAPST 763  
Db 1236 --HTEELVAAGTKTIKPCOSQSDPDVDTPTSTQKPKRSIRKADVEG---ELLACRNLMPSA 1292  
Qy 764 SPSLAVPTKPAEDSVETQVNDST-----SAETAQOM-----DV 798  
Db 1293 GKAMHTPKPSVGEEDKIIIFVGTVPQKLDLTENLTGSKRRRQTPKEEAQALEDLTGFKEL 1352  
Qy 799 DQEHSAEGSV-----CDPPATKADSVDEVVRPE----- 830  
Db 1353 FQTPGHTTEAIVAAGTKTKWPCSSPESADTPTSTRQPKTPLEKRDVQKELSALKLKTQ 1412  
Qy 831 -----NHASKYVEGDNTERKDLDR--ASEKVEPRDEDLVAQOINAQRPESQDNDSSATC 883  
Db 1413 TSGETHHTDKVPGGEDKKSINAFRETAOKLDP-----AASVTGSKRHPKTKKEAQLP- 1464  
Qy 884 SADEDVGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNPLDLLOH----- 932  
Db 1465 ---EDLAGMKELFQTPVCTDKPTTHEKTTKIACRS--QPDVDTPTSSKPSKRSRLKVD 1519  
Qy 933 -----RAAVTPMVVSCPTPCNIP--IGTPVSGVYALQ-----RHIKAMH 968  
Db 1520 VEEEFALRRKRTSPSAGKAMHTPKPAVSVEKNYAFMGTPVQKLDLTENLTGSKRRRLQTPK 1579  
Qy 969 ESAL-----LEQRQOEQIDLECRSS----- 990  
Db 1580 EKAQALEDLAGFKELFQTPRGHTESMTWNDKTAKVACKSSQPDLDKNPASSKRLKTSIGK 1639  
Qy 991 -----TSPCGTSKSPNREVEVLQAPAHQI-----TNLPEQVR 1023  
Db 1640 VGVKEELLAVGKLTQSGETHHTHTPEPGDGS---MKAFESPQKILDSAAASLTGSKR 1695  
Qy 1024 LPTTRTPRPPPLIPSSKTTTAVASKPSPIMGSGISQGTPTGLYLSHNOASTQOE--TPKPS 1082  
Db 1696 QLRT-----PKGSEVPEDLAGFI---ELFQ--TP-----SHTKESMTNEKTTKVS 1736  
Qy 1083 -----VGSISLGLPROESAKSATLPIYIKOEEFSPRSONSQEPGLLVRAQHEGVVRG 1134  
Db 1737 YRASOPDLVDTPSTSKPKQPKRSRKAD-----TEEEFLAFKQTPSAG---KAMH--TPK 1787

Qy 1135 TAGAIOESITRGTPTKSISVESIPSLRGSIQTGTPALPQTG--IPTALVKGSISSMPIE 1193  
Db 1788 AVGEKINDINTLGTBPVK-----LDQPGNLP-----GSRRL----- 1819  
Qy 1194 DSSPEKGREAAAGHVIYEGKSGHILSYDNKINARETRSPRTAHEISLRSKSYESVGN 1253  
Db 1820 ---QTRKEKAQ-----ALEELTGRELFPCTDNPDADEKTKKI--- 1857  
Qy 1254 IKOGMSMRSPVSAPLEGLICRALPRGS---PHSDLKERTVLSGSIOMGCTPRATESEFD 1310  
Db 1858 -----LCKSPQSPAD-----TPTNTKORPKRSLKADVEEFL---AFRKLTPSAGK 1902  
Qy 1311 GLKYPKQIKRSPPIRAPEGAIT-----KGKPYDGIITIKEMGR 1349  
Db 1903 AMHTPKAAVGEKIDINTFVGTVPVEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKELFQ 1962  
Qy 1350 S---IHEIPRODIITQESRKTPEVVQSTRPIIEGSIQGTPIKFDNNSGQ-----SAIKHN 1402  
Db 1963 TPGHTESMTDDKITEVSKSP-----QDPVKPTPTSSKQRLKISLGKVG 2007  
Qy 1403 VKSLITGPSKLSRGMPLLEIIVPENIKVVERGKYEDVKAGETVRSRHTSVVSGSVLRST 1462  
Db 2008 VKEEVLVPVKLUTQ-----TSGKTTQT--HRETAGDGKSI---KA 2041  
Qy 1463 LHEAPKAQLSPGIYDDTSAR--RTPVSYQNTMS--RGSPMMNRTSDVTIPPKNKSTNHERRK 1518  
Db 2042 FKESAKQMLDPANYGTGMERWPRTPKEEAQSLDLAGFKELFQTPD-----HTEE 2091  
Qy 1519 STLPTQRESIPAKSPVP--GVDPVVS--HSPFDPHHRGSTAGEV-----YWSHLP 1565  
Db 2092 ST--TDDKTTKIACKSPSPESMDTPTSTRRPKTPGLGRDIVEELSALKQLQTTHTDKVP 2150  
Qy 1566 TQLOPAMPFHR---ALDPAAAAVLFQRLSPTPGYSQ--YOLYAMENTROT--ILNDY 1617  
Db 2151 GDEKGINVFRETAQKLDLPAASVTGSKRQ--PRTPKKAQPLEDLAGFKELFQTPVCTDK 2209  
Qy 1618 ITSQOMQVNLDPDVARGLSPREQPLGPY---PATRGI-----IDLTNMPPTI-- 1662  
Db 2210 PTTHEKTKI---ACR--SQPDVPVGTPTIFKPSKRSRLKADVEEESALRKRTPSVGK 2264  
Qy 1663 --LVPHPGGTSPTPMDRITYIPTGQITFPPRPYNSASNSPGHPHLLAAASAEERERER 1720  
Db 2265 AMDTPPKPAGGDEKDMKFMGTPVQKLDLP-----GNLPG-----SKRWPTPK 2307  
Qy 1721 EKERERERIAAASDLYLRSGEQPRPGSHGYV--RSPSPSVRTQETMLOQSVFQGT 1778  
Db 2308 EKAQALEDL--AGFKELFQTPGTDKPTTDEKTTKACKSPQDPDVTPTASTQKPK----- 2361  
Qy 1779 NGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPA-----SRYNATAADALAALVDAASAP 1832  
Db 2362 ---RNLRRKADVEEFLALRRKRTPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGNLP 2418  
Qy 1833 QMDYSKTKESHEARLEENLRS-----RSAVSEQOOLEOKTLEVERKRSVQCLYTSSAF 1887  
Db 2419 --GSKRQPTPKKEALEDLVGFKELFQTPGHTESMTDDKITEVSKS----- 2466  
Qy 1888 PSGKPOPHS--SVVTSEAGKDKGPPPKSYEE---LTRGKTTITAAINFIDVIITR 1939  
Db 2467 ---PQSEFTSSSKQRLKIPLVKVDKMEPLAVSKLRTSGETTQTHTE-----PTG 2517  
Qy 1940 QIASDKDARERGSSSSSSSLSSHR-----YETPSDAIEVIS 1977  
Db 2518 DSKSIKAFKESPKQILDPAASVTGSRRLRTRKEKARALEDLVDFKELFSAPGHTESMT 2577  
Qy 1978 -----PASSPAP-----POEKLOTYQPEVVKANQANDPTROYEGPL 2014  
Db 2578 IDKNTKIPCKSPPPDELTDATSTKRCPKTRKPKRKEVKEELSVERLTQTSOGSTHKEPA 2637  
Qy 2015 -----HHVRPOQESPSQOOLPPSSOAGCMGOVPRTHRLITLITADHCQIIT 2060  
Db 2638 SGDRGKIVKORAKKKPNPVEEFSRRRPAPKEKAQPLEDLAGFKELTSETSGHTQESLT 2697  
Qy 2061 QDFARNQVSSQTPQOQPTST-----FQNSPSALVSTPVTRKTSNR--YSPE 2104



```

Db 2698 ASKATKIPCEPPLVDDTTASTKRHLRTRVQVQVKEEPSAVKFTQTSGETTDADKEPA 2757
Qy 2105 SQAQSVHQRGSRVSNLENLVDKSRGSRGKSPERSHVSSEPIPSPPQVPPVVEHKODS 2164
Db 2758 GEDGKIKALKESAKQTAPASVTVTSR--RRPRAPRESAQAIEDLAGFKDPAAGHTEES 2814
Qy 2165 LLLLSQGAEPAEQRNDARSPGISYLPSPFTKLENTS-----PMVSKKOEIIFRKLNS 2218
Db 2815 M-----TDDKTKIPCKSS-----PELEDATATSKRRPRTRAQKVEVKEELLA 2857
Qy 2219 SG-----GGSDMAAQP-----GTEIFNLPA---VTTSGSVSRGHSFADPASNGLLED 2265
Db 2858 VGTKUTQSGETHTKDKEPVGEGKGTAKQKAPAKRNVDAEDVIGSRROPAPRKEAQPLED 2917
Qy 2266 IIRKALMGSPDKVEDHGVGVMQPMGVVPGTAN---TSVVTSGTTRREGDPSPHSGVGC 2322
Db 2918 -----LASFOE-----LSQTPGHTEELANGAADSFTSAPKQTPDPSGAPLKISRVL 2963
Qy 2323 K-PKLISKNSRKSKSPGPGGYLGTERP 2350
Db 2964 RAPKVEPVGVVSTRDPVKQSQKSNSTSLP 2992

RESULT 14
YCS3_YEAST
ID YCS3_YEAST STANDARD; PRT; 1226 AA.
AC P25357;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE HYPOHETICAL 138.5 KDA PROTEIN IN RPS14A-GNS1 INTERGENIC REGION.
GN YCR033W OR YCR33W OR YCR592.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133166; PubMed=1776366;
RA Wicksstead B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
RT "The complete sequence of the unit YCR59, situated between CRY1 and
RL MAT, reveals two long open reading frames, which cover 91% of the
RP 10.1 kb segment."
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335897; PubMed=1872032;
RA Jia Y., Slonimski P.P., Herbert C.J.;
RT "The complete sequence of the unit YCR59, situated between CRY1 and
RL MAT, reveals two long open reading frames, which cover 91% of the
RP 10.1 kb segment."
CC -1- MISCELLANEOUS: THIS PROTEIN IS ENCODED BY A NON-ESSENTIAL GENE.
CC -1- SIMILARITY: SOME, TO S.POMBE SPAC2E12.01.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59075; CAA41799.1; -
CC EMBL; X59720; CAA42300.1; -
CC PIR; S15053; S15053.
CC PIR; S19445; S19445.
CC SGD; S0000629; YCR033W.
CC InterPro; IPR001005; -.
CC Pfam; PF00249; myb-DNA-binding; 2.
CC PROSITE; PS50090; MYB_3; 1.
CC Hypothetical protein.
CC DOMAIN 594 599 POLY-SER.

```

```

VARIANT 305 305 T -> A.
VARIANT 375 375 N -> S.
VARIANT 404 404 H -> Q.
VARIANT 435 435 A -> V.
VARIANT 442 442 D -> N.
VARIANT 644 644 Y -> H.
VARIANT 801 801 E -> D.
VARIANT 806 806 K -> Q.
VARIANT 831 831 I -> V.
VARIANT 846 846 E -> G.
VARIANT 851 851 A -> D.
VARIANT 855 857 GVR -> AVQ.
SQ SEQUENCE 1226 AA; 138503 MW; 8D133A0918658E53 CRC64;

Query Match 3.1%; Score 390; DB 1; Length 1226;
Best Local Similarity 20.2%; Pred. No. 8.6e-08;
Matches 222; Conservative 181; Mismatches 368; Indels 326; Gaps 52;

Qy 31 NTRHQQFAVPDYRSSHLEVSQAOLQOQQOQLRRRPSL--LSEFHFGS--DRPQ----- 83
Db 247 NSHQRE--PFWKAN-----STILKSTHSQS---SPSLHTKKFHDANKLDKPEASVK 294
Qy 84 -----ERRT-SYE-----PFHPG----- 95
Db 295 VETPSKDETKTISYHDNNFPKRKSVSKNPAPLEPDNIKVGEDALGKKEVHKGSGREIAKE 354
Qy 96 -PSVP---DHDSLESKRPRLEQVSDSHFQVSAAVLPLVHPLPEGLRASADAKKDPFEG 151
Db 355 HPTVPVKHEDELEAKKVNKI-----NIDGKQDEIWT 389
Qy 152 KHEAPSPISGQPCDDQNASPSKLSKELIQSDMRVDR-EIAKVEQ---QILKLKKKQ 206
Db 390 AKTVASA-----VEVSKEKHELTRSVKESPEIRYERAYDKALKTATK 437
Qy 207 QOLEEAAKPPPEP-----EKVPSPPPVEQKHSIYQIYDENRKK--ABEAHKIFEGLPK 260
Db 438 LTVDDDNKSYEEPLEKVEGCIFFPLP---KAETRLWELKNQKRNKIISKQYLLKKAIRNF 494
Qy 261 VELPLYNQPSDTKVYHENIKTNQVNRKKLILFFRRNRHARKOREQKICQVDOLMEANEK 320
Db 495 SEYFFYQAQ---NKLHQ--QATGLITLTKIISKIRKEHLKK---INLKHDFDLOKKYK 546
Qy 321 KVD---RIENPRRKAKESKTREYVEKQFPEIRKOREQOERQFQVQGA--GLSATIAR 375
Db 547 ECEILTLSLENLKEETENKKEH-----ELMEOKRREEGIEETEKEKSLRHPSSSSSR 600
Qy 376 SEHEISEIIDGLSQEN-----NEKOMQLSVIPPMFDAEQR-RVKFINMNGLMED- 426
Db 601 RRNRADFVDD--AEMENVLLQIDPNYKHYQAAATPPLIDPIRKYSYKFCDDVNNLVTDK 658
Qy 427 ---PMKYVKDRQPMNVVTDHKEIFKDKFIOHPKNFGLIASYL-ERKSVPCVLYIYLT 482
Db 659 KLWASRLKDS--DNFTDHEHSILFEGYLTHPKKFKGISHYMGGLRSPEECVLHYRTK 716
Qy 483 KNEYKALVRRNYGKRGRNQOIARPSOEKVEKEEDKAETKEKEKK--DEEEXDE 540
Db 717 KTVYKQLLIDKNKKR-----MSAAAKRRKKERSNDEVEVEDESKESTTIDKEESE 772
Qy 541 KEDSKEN-----TKEKDIDGTAEETEREQATPRGRKTAN--SQGRKRGRITRSM 590
Db 773 -NNAEENQPVLVQSGVEKGDPLGTPEKVE---NMIEKRGEEFAGELENAERVNDLKAHD 829
Qy 591 NEAAAASAAAAATEEPPLPPP-----PEPI-----STEPVET 625
Db 830 EIGESNKS SVIETNNEVQIMAPKGGVRNGYVPEETKELDFSLLENALQKKHKSAPHEKT 889
Qy 626 SRWTEEMEYAKGLVEHGRNWAIAKVMGKTSQAQCKNFYFNKRRHNLNLLQOQKQK 685
Db 890 SYWSVRESQPLPELLKEFGSOWSLISEKLGHKSTWVRYNYQORNAARNGWKLIVDETDLK 949
Qy 686 TSRKPREERDVQSCEVASTVSAQEDIEDIASNEEENPDESEFAVK--PSEDSPE----- 739

```



Db 950 -----RDGTSSES-----QQSQILIPERNINAYSNIIPPQORPALGYF 989  
Qy 740 -----NATSGNTEPAVE-----LEPTETAPSTSPSLAVPSTKPAED- 777  
Db 990 VQOPHTGHNTSISSIDGSRIFGPDHFDHTSKISAPLTTLPPLPSIQPPRSEMAEPT 1049  
Qy 778 -----ESVETQVNDISIAETAQMDVDQOQHSASEGVCDDPPATKADSVDEVRV 828  
Db 1050 VTDLRNRLDHDIT- LADAASSVT-----NNQNFSENRNAI----- 1084  
Qy 829 PENHASKVEGNTYKRLDRASEKVEPRDEDLVVAQINAORPEPQSDNDSSATCSADE- 887  
Db 1085 -----DIGRKSSTIS-----NLLNDRSMKSKSFQASRHEA 1116  
Qy 888 DVDGEPRQRMPMDKSPKLNLPSTILVSSPLKP-----NPLDLPOL--QHRAAVIP 938  
Db 1117 QLETPSNMNIIVGQIKRNTTPRSS-SISALLAPVNGQSDNPGRLPLPFQHAISQGT 1175  
Qy 939 PMVSCPTPCNIPGRPV 955  
Db 1176 PTF---PLPAPRTSPIS 1189

RESULT 15  
PGCV\_CHICK  
AC Q90953; Q90945; STANDARD; PRT; 3562 AA.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).  
GN CP5G2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN-WHITE LEGHORN; TISSUE=Limb bud;  
RA MEDLINE=93300846; Pubmed=9314802;  
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan  
RT expressed during chondrogenesis in chick limb buds. Alternative  
RT spliced multiforms of PG-M and their relationships to versican.";  
RL J Biol. Chem. 268:14461-14469(1993).  
CC -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN  
CC CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN  
CC THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS  
CC HYALURONIC ACID.  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; V0 (SHOWN HERE) AND  
CC V1; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: PRECHONDROGENIC CONDENSATION AREA OF  
CC DEVELOPING LIMB BUDS.  
CC -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT  
CC (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X60226; CAA42787.1; -

DR EMBL; D13542; BAA02742.1; -  
DR HSP: P00740; IIXA.  
DR InterPro; IPR000152; -  
DR InterPro; IPR000436; -  
DR InterPro; IPR000538; -  
DR InterPro; IPR000561; -  
DR InterPro; IPR001304; -  
DR InterPro; IPR001881; -  
DR InterPro; IPR003006; -  
DR Pfam; PF00008; EGF\_2;  
DR Pfam; PF00193; Xlink; 2.  
DR Pfam; PF00047; ig; 1.  
DR Pfam; PF00059; lectin\_C; 1.  
DR Pfam; PF00084; sushi; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
DR PROSITE; PS01241; LINK; 2.  
DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
KW Hyaluronic acid; Alternative splicing.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 3562 VERSICAN CORE PROTEIN.  
FT DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 166 243 LINK 1.  
FT DOMAIN 264 345 LINK 2.  
FT DOMAIN 3254 3290 EGF-LIKE 1.  
FT DOMAIN 3292 3328 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 3330 3459 C-TYPE LECTIN.  
FT DOMAIN 3460 3518 SUSHI.  
FT DISULFID 44 129 BY SIMILARITY.  
FT DISULFID 171 242 BY SIMILARITY.  
FT DISULFID 195 216 BY SIMILARITY.  
FT DISULFID 269 344 BY SIMILARITY.  
FT DISULFID 293 314 BY SIMILARITY.  
FT DISULFID 3258 3269 BY SIMILARITY.  
FT DISULFID 3263 3278 BY SIMILARITY.  
FT DISULFID 3280 3289 BY SIMILARITY.  
FT DISULFID 3296 3307 BY SIMILARITY.  
FT DISULFID 3301 3316 BY SIMILARITY.  
FT DISULFID 3318 3327 BY SIMILARITY.  
FT DISULFID 3334 3345 BY SIMILARITY.  
FT DISULFID 3362 3454 BY SIMILARITY.  
FT DISULFID 3430 3446 BY SIMILARITY.  
FT DISULFID 3461 3504 BY SIMILARITY.  
FT DISULFID 3490 3517 BY SIMILARITY.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 709 709 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 948 948 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. .) (POTENTIAL).



FT. VARSPLIC 485 1411 MISSING (IN ISOFORM V1).  
SQ. SEQUENCE 3562 AA; 388078 MW; 9BC566E88C1602D2 CRC64;

Query Match 3.08; Score 374.5; DB 1; Length 3562;  
Best Local Similarity 18.18; Pred. No. 1.1e-06;  
Matches 511; Conservative 379; Mismatches 1102; Indels 833; Gaps 118;

QY 9 NOGAFSTEQ-SRYPHSHVQ-----YTFNTRHOQ--EFAVPDVRSSHLEVSQASQL 56  
Db 781 SEGRFSGEKTHPPVSGMLQTDKQOYYTEETSHKRIELEDDEDSMGEPSSPGQI 840  
QY 57 LQ-----QQOQOQLRRRPLLSEFHGSDRPOERRTSYFFHPGPPS----- 97  
Db 841 IEVTKHLGAPVSAVTDKTSMTAETESDEEVVSADFDQTKGT-BVFHTSSLDLEKF 899  
QY 98 -----PVDHDS-----LESKRPLEQVSDSHFORVSAAVLPLVHPLPEGLRA 139  
Db 900 TLSKIPEDSSATVKSFSSSGTVLPTAVATVLEVTHEADETSGYVLNMFTSTPEGEOR 959  
QY 140 SADAKKDPATFGGHEAPSSPISGPGDDQNASPSKLSKEELIQSDMDRVOREIAKVEQOI 199  
Db 960 KA-TEKSPA-----TSADEVSTGCTESKTMTEG-QGISVTSAAKESV 1002  
QY 200 LKAKKKOQO-----LEEAAKPPPEKPVSPPPVE-----QKH 232  
Db 1003 AALQEREQPSVGLPETKEPKFTDVTETIETVTPQREGDSLVPVTVGSEDIGEMQVTDH 1062  
QY 233 RSIVOOIYDENRKAEAHAKIFGLGPKVBLPLYNQSDTKVYHEN---TKNQVMRKL 289  
Db 1063 TSFDSIIHTAATVTSKASEVF---PK-ELSTKDQDRELGTAMGSTLPVTSVQMHEQKT 1117  
QY 290 ILFKRRNHARKOEOKICORYDQLEMAWKKVVDRI---ENNRPRKAKESKTRVEYEQF 346  
Db 1118 TAGFESQTTQEKHDENGAYDENYDPAELSVPALMLTEYQVSGVPEVTSRSLHTGT 1177  
QY 347 PEIRKOREQOQRFVORGAGLSATIAARSEISEIIDLSEQOENNEKQMLSVIPPM 406  
Db 1178 PKAETATDQEEKITEAVPVTFGTQAKVYESK-----GTTTREDRDVGWNSVLP 1229  
QY 407 MFDAEQRRVFINNGLMEDPMKYKDRQPMNVWTDHEKEIFDKDFIQHPKNGLIASYL 466  
Db 1230 TMLSSPSTAGSISLLTLGASPS-----OTPEGSGISEELE 1264  
QY 467 ERKSPVPCVLYLYLTKKNENYKALVRNRYGKRGRNOQIARPSQOEKVEEKEDEKAE 526  
Db 1265 EVKIVP-----FSSRATDKTVI-----SDLTSSISAVDKIQTSASKPVSS 1308  
QY 527 KKEEKKDEEKEKED-----SKENTKEKDKIDGTAETEEREQATPRGRKKTANSQ 578  
Db 1309 KSPRIIPEDEEVTSSDIIVIDESISPSKASAEDDLTKMVEPE----- 1352  
QY 579 GRRKGRITRSMTEAAAAAATAEPPPLPPPEPTEPETSRTWTEEMEYAKK 638  
Db 1353 -----IDKEYFTSSATAVAPTPTVNEATEALQOPEVSPSTSHPOSGTDIR-- 1400  
QY 639 GLVEHGRNAAIAKMWGTSKSEAQCNFYFKRRHNDLNLQOHKQTSRKPREERDVQS 698  
Db 1401 -----LVVIQITGNDTDHPVNEFLDLFSRH-----ILPHAVDETHDAESAQTEP 1445  
QY 699 CESVASTVSA-----QEDIEDIE-----ASN 718  
Db 1446 CTSDSVDSSESJILDPFFPFMFDFEEDDECENTDVTTPPALQFNGKQVTSAPKST 1505  
QY 719 EENPEDESEVAVKPSDPSNATSRGNETPAVELEPTTETAPS-----TSPSL 767  
Db 1506 KAEARSDQIESVAHSKNVTFSQINETNTIIEETASGTMQPSKAGVAGAEVETQPTA 1565  
QY 768 AVPSTKPAEDESVEVQVNDISAEFTA---EQMDVDQOE---HSAEESGVCDPPPPATKADS 821  
Db 1566 DVAMLEPVYSESEVTTTDKYLEITSVYEQSPKKNKETVMWHGTEESSTKD---TK--- 1618  
QY 822 VDVEVRVPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINAQRPESQDNDSSA 881

Db 1619 ---NLLLTNESS---GDGSTESDLRS-----VFTEI-----L 1646  
QY 882 TCSADEVDGEPEPERQRMFPMDSKPSLLNPTGSGILVSSPLKPNP-----LDLPOLQHRANV 936  
Db 1647 TMSHED-----SEKISHTSVPTILSVERSAVTAAPSADSDTATGIDVKDL----- 1694  
QY 937 IPPMVSTPCNIPITGTPVSGVALYORHIKAMHESALLEEQROEQI----- 983  
Db 1695 IPKGGTATPGN-----YKSTIKLDAEPPFSPSNPEATSHHTKPKDMTASSFVLE 1743  
QY 984 ---DLECRSTSPGCTSKSPNREWEVLQAPAPHQLITNLPEGVRLPTTRTPRPPPLIPSS 1040  
Db 1744 GSGDVEENSTLASAMTTETAVALTSLVQD-----TSLGSGTVLPTEIS-----VTIS 1790  
QY 1041 KTVIVASKEPFIIMGSGISOGTPGTYILSHNOASVTOETPKPSVGSISLGLPQOESAKSA 1100  
Db 1791 EITPALPGGTRILYSTFDQSEATVST-----NFVSELIMEVQVGVSSVATEKKVEDEKEV 1845  
QY 1101 -TLPYIKOE-----EFSP-----RSONSOPEGL-----LVRAQHEGVVGTGA 1136  
Db 1846 QTVVYSQEIISTTDKAKKSELDERGTTNEVRTVTSQBPPLREIVPITGTMHSEIKKVTGA 1905  
QY 1137 GA-----IQEGS-----ITRGTPTSKISVESIP---SLRGSITQGTALPOTGIPT 1180  
Db 1906 TPLREKLFINEGSAEPPADLFAAGSPTRKVVSTDSPTDSSGDDIDVITESATLTSVPSR 1965  
QY 1181 ALV-----KGSISRMPIEDSSPEKGBREAAKSHVYIEGKSHILLSYDNKKNAREGTR 1233  
Db 1966 SVIETQTVKHEGNINVISLKNTTTEYEE-----HI---GTGGPVTSVSS-----TGSD 2012  
QY 1234 SPRTAHSILKRSVESYEGNIKQGMKMSRSPVSPLEGLICRALPRGSPHSDLKERTVLS 1293  
Db 2013 GLTEESEVAITEMS---ENVFSTENQGEPTQE-----AVPTTAPSDIASKRL--- 2055  
QY 1294 GSIMQGTPTATTESFEDGLKYPKQIKRESPPIRAFEGAITKGKPYDGTITTIKEMGRSIEH 1353  
Db 2056 GSRREVTSHT-----PVIRT----- 2071  
QY 1354 IPRQDILITQSRKTPPEVQVQSTRPIIEGSIQGTPIKFDNNSGQSAIKHNKSLITGSKL 1413  
Db 2072 ---KOLETAETVSSPESV-----VNNSLDITMVTHGTIRAVAESST 2110  
QY 1414 SRGMPPIEIVP-ENIKVVERGKYEDKAGETVRSRHTSVVSSGSPS---VLRS----- 1461  
Db 2111 KKGKGSFSAVSLGLKLMIEHSGEELK---VDSSTTKLMSNGPTEKLGSHSFDDGS 2166  
QY 1462 ---TLHEA--PKAQLSPGIYDDTSARTTPVSY-ONTMSRSGSPMMNRSDVTIPPNK---S 1512  
Db 2167 GEATLETSETKASVSP-----TGKPEPQEQYGRKTVSMPSAVVHA---YTAEPNELVTS 2218  
QY 1513 TNHERKS--TLTPTQRESIPAKSPVPGVDPVVSHPF--DPHHRGSTAGEVYWSHLPTOL 1568  
Db 2219 TEHDITSLQVITDTEMEKKAANELT---VTSFATNLPLSEDVHS-----WEDRPREI 2267  
QY 1569 DPAMPFFRALDPAAAA---YLFQRLS-----PTFGPQYQOYLAMENTRQTL- 1614  
Db 2268 LP-----KAJESSCEATEDPFFISTQANHEHVEFLSVPTIRPHSEENKVEAESDEKILLP 2322  
QY 1615 --NDYIT-----SQOMQVNLRP--DVAR-GLSPREQLGLPYPATR 1650  
Db 2323 FNNDRVTESAVIERKYLSSPFTDTEQEELVQNIFFTEDIPRLFLTKPEE----- 2372  
QY 1651 GIIDLTNMPT--ILVPHPGTSTPPMDRITYIPG-----TQITFPPRPYNSASMP 1700  
Db 2373 -----KPTNNELISDPLFSGOGSGDEFTVIPSVESLAVKETTNTLSPPFPFASVGP 2424  
QY 1701 ---GHPTHLAAAAAERERE--REREKERERERERIAAASDLYLRPGSQBPGRPGSHGVRS 1756  
Db 2425 KLSDTKQVFEESGSDSNABEINEIRITTTAAELTETAYSMATSSPALEEESSHSNSDKD 2484  
QY 1757 PPSVTRTQETMLQORPSPVFOGTNGTSTVITPLDPTAQLRIMPLPAGGSPISOGLPASRYNT 1816



Db 2485 ITHYFLVIEDPYNKEMDRHRRGENGTS-----RPLTPG---DVSLESSHML 2528  
Qy 1817 AADALAALVDAASAPOMDVSK---TKESKHEAARLEENLRSA--AVSEQ-----Q 1864  
Db 2529 TDDVTPVSVILSETPYLEMGCKSLATSNKMPRSRVLPSSGEGSGWGVDSFAPDILTH 2588  
Qy 1865 QLEQKTLVEKRSVQCLYTSSAFFSGPKPQPHSSVVSAGKDKGPPPKSRYEELTRGK 1924  
Db 2589 STAFSVMEVE-----LTASSHIPG-----VSEVMTTHVPGDGS-----Q 2623  
Qy 1925 TTIT--AANFTD--VIITRQIASDKDARERGSSQSSSSLS-----1962  
Db 2624 TVITGLASLFTFEEREIVANRTAADPKTGTSELTSDTGMSLDIIPVDDRRHVTLNVSU 2683  
Qy 1963 -----SHRYTPSDATEVISPASSAPPOE--KIQYQPVVVKANQAEADPTR---QY 2010  
Db 2684 GDYFLIERLQIPSEKTIIDMDHSMKMPEDIIISVQTPNPNLVIRSTQVSDDNMKAEDKY 2743  
Qy 2011 EGPLHHRPQOES-----PSPQOQLPPSSQAEGMGQVPRTHRLITLADHICQITQDFA 2064  
Db 2744 DSIINFSTVEENSFGDNLSTTSIQPSSSESVTAGHGPKLVDRKDLGSGYAMQFATETLT 2803  
Qy 2065 R-----NOVSSOTPOOPTSTFONSPSALVSTPVRTKTSNRYSPESQASQV 2110  
Db 2804 TTVLNEGLIFLTPVPSLVSPHMPHESKESEFEAKHIGRTST-----TDDVVEPYTSANN- 2857  
Qy 2111 HHQPGSVSPENLVDKSR---GSRPCKSPERSHVSSEPYEPISPPQVPPVHEKQ---- 2162  
Db 2858 -----QVITDQSKTMSISFGMGQGESEG-DKKPMIPSLTPDLTMTETKALTDD 2905  
Qy 2163 --DSLLLLSQGAEPAE-QRNDARSPGSIYL--PSFFTKLENTSPMVKKQKQIFRKLN 2217  
Db 2906 TFDVSMVTQSMQSHATVSSSSSEKSHVTYMQTKASTEYEEDSV-----SLN 2955  
Qy 2218 SSGGSDMAAQAQGTGFNLPATVTSVSGSVSRGHSFADPNASNLGLEDIIRKALMGSD 2277  
Db 2956 SVSQNPSSVTVLWVNGVSKYPIVITPSTSAKDSQSDSHSD-----GTFKE 3003  
Qy 2278 KVEDGVVMSOPMGVPTGANTSVVTGSTRRE--EGDPSPHSGVCKPKLISKNSRKS 2335  
Db 3004 VSSDMAATYKPTDLDITVTSLLVFPPESEISISTEPH-----FNKFTVTERS 3054  
Qy 2336 KSPIPGOGYLGTERPSSV-----SVHSEGDYHROTPGAWEDRPS-----STGSTOFF- 2384  
Db 3055 EETESSVNDLIIENATVSGDSPSIH--DY--PTAFWNFGERTSDVPKLSLTIEVEFS 3109  
Qy 2385 ---YNP-----LTWMLSTPTTPPIACAPSAVNAQA--PHQONRIWEREPAPLLSAYE 2433  
Db 3110 ERVKNPQESDRSTERERPRLSAPVSDSPNSIEVGFKPDQ-----EAVTMTLSLE 3162  
Qy 2434 TLDSD 2438  
Db 3163 PLDRS 3167  
RESULT 16  
ID APC\_HUMAN STANDARD; PRT; 2843 AA.  
AC P25054; Q15162; Q15163;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).  
GN APC OR DP2.5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91335210; PubMed=1651563;  
RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,  
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,  
Finnear R., Markham A., Groffen J., Boquski M.S., Altschul S.F.,  
Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.;  
"Identification of FAP locus genes from chromosome 5q21."; Science 253:661-665(1991).  
[2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Fetal brain;  
MEDLINE=91330307; PubMed=1678319;  
RA Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,  
Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,  
Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,  
Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,  
Abderrahim H., Cohen D., Leppert M., White R.;  
"Identification of deletion mutations and three new genes at the  
familial polyposis locus."; Cell 66:601-613(1991).  
[3]  
RP ASSOCIATION WITH CATENINS.  
MEDLINE=94082295; PubMed=8259519;  
RA Su L.-K., Vogelstein B., Kinzler K.W.;  
"Association of the APC tumor suppressor protein with catenins."; Science 262:1734-1737(1993).  
[4]  
RP REVIEW ON VARIANTS.  
MEDLINE=94154728; PubMed=8111410;  
RA Nagase H., Nakamura Y.;  
"Mutations of the APC (adenomatous polyposis coli) gene."; Hum. Mutat. 2:425-434(1993).  
[5]  
RP VARIANTS FAP.  
MEDLINE=91335211; PubMed=1651563;  
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,  
Koyana K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,  
Petersen G., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,  
Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;  
"Mutations of chromosome 5q21 genes in FAP and colorectal cancer  
patients."; Science 253:665-669(1991).  
[6]  
RP VARIANTS FAP.  
MEDLINE=93285030; PubMed=1338904;  
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,  
Miki Y., Mori T., Nakamura Y.;  
"Somatic mutations of the APC gene in colorectal tumors: mutation  
cluster region in the APC gene"; Hum. Mol. Genet. 1:229-233(1992).  
[7]  
RP VARIANTS FAP.  
MEDLINE=93244793; PubMed=1338691;  
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,  
Nakamura Y., Horii A.;  
"Somatic mutation of the APC gene in gastric cancer: frequent  
mutations in very well differentiated adenocarcinoma and signet-ring  
cell carcinoma."; Hum. Mol. Genet. 1:559-563(1992).  
[8]  
RP VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 & S-2502.  
MEDLINE=93250848; PubMed=1338764;  
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,  
Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,  
Baba S., Nakamura Y.;  
"Screening for germ-line mutations in familial adenomatous polyposis  
patients: 61 new patients and a summary of 150 unrelated patients."; Hum. Mutat. 1:467-473(1992).  
[9]  
RP VARIANT FAP TRP-99.  
TISSUE=Peripheral blood lymphocytes;  
MEDLINE=95134544; PubMed=7833149;  
RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Roth J.,  
Mueller A., Mueller H., Scott R.J.;  
"Mutational analysis of the first 14 exons of the adenomatous  
polyposis coli (APC) gene."; Eur. J. Cancer 30A:1709-1713(1994).







Db 1371 SPPEHYVQETPLMFSRCTSV--SSLDSEFSRSIASSVQSPCGSMVSGILSPSLDLPDSPGQ 1429  
QY 741 A--TSRGNTPEPAVELETTETAPSTSLAVPSTKPAEDSVETQVNDSSIAETAQMDV 798  
Db 1430 TMPESRSKTPPP--PPQATQKREVPKNAPTAKRESGPKQAAVNAVO-RVQVLPDA 1485  
QY 799 DQOQSHSAEE--GSCV-----DPPATKADSDVDEVR-----VPEN-HASKVGGDN 840  
Db 1486 DTLHFATETPDGFCSSLSALSLEDPFQK-----DVELRIMPVQENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDELVVAQAQINQARPEPQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNEQKEAEKTI DSKDL-----DSD-----DDDEILEECIIS 1581  
QY 891 ---CEPERQRMFMDKPSLLNPTGSLVSPKPNPLD---LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSSKAKKPAQATKSLPPPVAR-----KPSLPVYKLLPSQNRLQPKQHVSTP 1634  
QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRYVCVEGTPINSTATSLDIEPSPNBLAAGEVGRGQAQGEFEKRTIPTEG 1694  
QY 971 AL-----LEEQROQIDLECRSSSTSPCGTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGKTSSTVITPELDDNKAEGLDILAEICINAMPKSKHKPFRVKKIMDQVOQAS 1754  
QY 1008 --QAPHLITNLEPGVRLPTTRTPRPPLI-----PS 1039  
Db 1755 ASSAPNPK---NQLDGGKKKPTSPVKIPQNTYRTRVRKNADSKNNLNAERVFSDNKDS 1811  
QY 1040 SKTTVAASEKPSF-----IMGGST-----QGTGPGYLTSHNOASYQTETPK 1080  
Db 1812 KKQNLKNSKDFNDKLNEDNRVGRGSAFDSPHHYTIEGTP--YCFSRND----- 1860  
QY 1081 PSVGSISLGLPROQESAKSLPYIKQEEFSPRSQNSOPEGLLVRAQHEGVYRTAGAIQ 1140  
Db 1861 -SLSSLDQDDVDLSREKAEKRAKENKESEAKVTSHTE--LTSNQOSA---NKTQAIA 1914  
QY 1141 EGSITRTGPTSKISVESIPSLRGSITGTPALPOTGIPTALVK-----GS 1186  
Db 1915 KQINRGOPKPILOKQS-----TFQSSKDIPRGAATDEKQNFALIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSPEKREAAKSGHVIYEGKSGHI----- 1219  
Db 1969 LSSLDIDQENNNKENEFIKETPPDSQGEPSKQASGAPKSFHVEDTPVCFSRNSLS 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRAHEISLKRSTVESVGNKQGMWRESVPVSAPE 1270  
Db 2029 SLSIDSEDDLQECISSAMPKPKKPSRLKGDNEKHSRNMGGILGDLTLDLKDQIRPDS 2088  
QY 1271 GLICRALPRGSPHDLKERTVLGSIIMOGTPRATTESFEDGLKYPKQIKRESPIRAFEG 1330  
Db 2089 E---HGLSPDSENFWDKAIQEGANSIVSSLUHQAAC-----LSRQASDSDSILSKS 2140  
QY 1331 AITGKPKYDGIITTIKEMGRSITHEIPRODILTOESRKTPVEVQVSTRPIEGSISQGTPIK 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPFTSNKGPRIKL-----PCKSPLETKKI 2181  
QY 1391 DNNS-GOSATKHNVKSLITGP-----SKLSRGM-PPLIEIVENTIKVBERGYEDVKAETV 1444  
Db 2182 ESESGKIGKKGVYKSLITKVRNSNSIQMKQPLQ---ANMPSISRG----- 2227  
QY 1445 RSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGSPPMNRSTD 1504  
Db 2228 -----RTWHI-----IPGV-----RNSSTSPVSKKGP 2252  
QY 1505 VTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDVSVSHSPDFPHHGRSTAGEYWS 1562  
Db 2253 LKTPASKSPSEGQATTSRPGAKSVKSELSPVARQTSQIGGSKAPSRSGSR----- 2305  
QY 1563 HLPQLDPAFPFHRALDPAAYLFRQLSFTPGYPSQYQLYAMENTRQTLNDYITSQ 1622  
Db 2306 -----DSTPSRPAQPL-----SRP 2320

QY 1623 MQYNLRPDVA---RGLSPREQPLGLPYPATRGIIIDLTNMPPTILVHPGCTSTPPMDRIT 1679  
Db 2321 IQSPGRNISIPGRNIIISPNNK-----LSOLPRT-SSPSTASTKSSGSKMS 2365  
QY 1680 YIPGTQITFPFPPVNSASMSGPHPTHLLAAASAAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQONLT--KOTGLSKNASSIPRESASAKGLNOMNGNGANKVYELS 2417  
QY 1728 RIAAASDLYLRPGSEQGRPGSHGYVRSPSPSVRTQETMLQQRPSPVFOGTNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSEPRVLVROSTFIKEAPS----- 2449  
QY 1788 DPTAQLRIMPLPAGGPPISQGLPASRYNTAADALAAALVDAASAPQMDVSKTKESKHEA 1847  
Db 2450 -PTLR-KLEESASFESLS---PSSRPASPTRSQAOPTVLSPSLPDMSL-THSSVQAGG 2503  
QY 1848 --RLEENLRKSAVSEQQLEOKTLEVEKRSVQCLYTSSAFFPSGKPKQPHSSVVYSEAGK 1905  
Db 2504 WRKLPPNL---SPTIENDGRPAKHDIAR-----SHSESPSLPINRSGTWKREHSK 2553  
QY 1906 DKGPPPKSYEEELTRGKTTITAN-----FIDVIITRQIASDKDARERGSQSDS 1957  
Db 2554 HSSSLP--RVSTWRRTGSSSILSASSSESEKAKSEKHNVSISGTKQSKENQVSAKGT 2611  
QY 1958 SSSLSHRYE-TPSDALEVISPASSPAPQOEKLOTYOPEVYKANAQENDPTROYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTSTSVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668  
QY 2017 YRQOESPSPOQLPP-----SSQAEQMGQVPRHRLITLADHIC 2056  
Db 2669 PR---SGRSTGNTPPVIDSVSEKANPNIKDKNDQAKQNVGNSVPM--RTVGLNRLN 2723  
QY 2057 QITQDFARNQVSSQTP-QQPTSTFONSPSALVS-TPVRTKTSNRYSPESQAQSVHHQR 2114  
Db 2724 SFIOVDAPDQKTEIKPQONNPVSETNESSIVERTPFSSSSSKHSSPSGTVA----- 2778  
QY 2115 PGRSVENLVDRSGRSGKPSRSHVSEPEPISPPOQVPPVH--EKODS 2164  
Db 2779 --AKVTFFNY-----NPSRKSDSTARSQIP-TPVNNYTKKRD 2818  
  
RESULT 17  
TPR\_HUMAN  
ID TPR\_HUMAN STANDARD; PRT; 2349 AA.  
AC P12270;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NUCLEOPROTEIN TPR.  
GN TPR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93064711; PubMed=1437155;  
RA Mitchell P.J., Cooper C.S.;  
RT "The human tpr gene encodes a protein of 2094 amino acids that has  
RT extensive coiled-coil regions and an acidic C-terminal domain.";  
RL Oncogene 7:2329-2333(1992).  
RN [2]  
RP REVISIONS, AND CHARACTERIZATION  
RX MEDLINE=95096166; PubMed=7798308;  
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,  
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;  
RT "Tpr, a large coiled coil protein whose amino terminus is involved in  
RT activation of oncogenic kinases, is localized to the cytoplasmic  
RT surface of the nuclear pore complex.";  
RL J. Cell Biol. 127:1515-1526(1994).  
RN [3]  
RP SEQUENCE OF 1-142 FROM N.A.



RX MEDLINE-89262257; PubMed-3387099;  
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;  
 RT "tpr homologues activate met and raf";  
 RL Oncogene 2:617-619(1988).  
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE  
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS  
 CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE  
 CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH  
 CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER  
 CC COMPONENTS, INCLUDING P62.  
 CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND  
 CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.  
 CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK  
 CC OR RAF GENES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X66397; CA47021.1; -;  
 DR EMBL; Y00672; CA68681.1; -;  
 DR PIR; S00928; S00928.  
 DR MIM; 189940; -;  
 KW Heptad repeat pattern: Coiled coil; Proto-oncogene;  
 KW Chromosomal translocation; Nuclear protein; Transport.  
 FT DOMAIN 78 360  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 422 571  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 575 628  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 758 805  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 834 869  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 934 979  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1004 1064  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1138 1166  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1196 1241  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1262 1304  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1354 1434  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 1476 1595  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 527 530  
 FT POLY-SER.  
 FT DOMAIN 1833 1836  
 FT POLY-GLU.  
 FT DOMAIN 1957 1964  
 FT POLY-ASP.  
 FT DOMAIN 2295 2298  
 FT POLY-ASP.  
 SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 2.89; Score 360; DB 1; Length 2349;

Best Local Similarity 18.39; Pred. No. 2.3e-06;

Matches 484; Conservative 394; Mismatches 966; Indels 806; Gaps 107;

QY 24 SVQTFNTRHQEQFAVDYRSSHLEVSQASQLQO-----QQOQLRRRPSLLSEPHGSD 80  
 DB 108 AIQOFTKEELEAEKRLDITNRLSQELEYLTEDYKRLNKESNTTKGELQLKLD 167  
 QY 81 RPOERTSYEPFPGSPVDHDSLESKRPRLEQVSDSHFQVRSAAVLPLVHPLPGLRAS 140  
 DB 168 ELQ-----ASDVSVKYREKLEQEKE-----LLHSQNTWLNTE 200  
 QY 141 ADAKKDPAGGKHEAPSPISQPCGDDQNASPKLSKEELIQSDMRVDREIAK-VEQOI 199  
 DB 201 LKTKTDELLALGREGKNEILEKCNLENKKEVSRL--EQOMGLKTSNEHLQKHVEDLL 258  
 QY 200 LKLK-KQQO-----LEEAAKPEPEKPPVPPVQKHRSIVOII 239  
 DB 259 TKLKEAKEQQAAMEEKFFHNELNHAKLSNLYKSAADSEAKSNELTRAVEELHKLL----- 314  
 QY 240 YDENRKAEEAHKIFEGIGPKVELPLYNQPSDTKYVHENIKTNQVMRKL-----ILFFK 294  
 DB 315 -----KEAGANKAIQDHLLVE-----QSKDQMEKEMLEKIGRLEKELEANDLLSAT 363

QY 295 RRNHARKREQ-----KICORYDQIMEAWK-KVDRIENRRRAK 334  
 DB 364 KRKGAILSEELAAMSPTAAAVAKIVKPGMKLTLEYNAVYVETQDQLLEKLENKRINYL 423  
 QY 335 ESKTREYKEKOPFETRKOREQOORFQVQAGLISATIAEISEHSEIIDLGSQENNE 394  
 DB 424 DEIVKE-YEAKAPILKROREYERAKA---VASLVSLEQAMKEIQLR-----QEDTD 473  
 QY 395 KOMRLSV-----IPPMFADABOR----- 414  
 DB 474 KANKOSSVLERDRNRMEIQVKDLSQLIRVLMELEAARGNHVIRDEEVSSADISSSEVI 533  
 QY 415 ----VKFINMGLMDPMKYKD-RQFMNVWTDHEKFKDKFIQHPKNFGLIASYLE-- 467  
 DB 534 SQHLYSYRNIIELOOQONRLVALRELGETREREQETSSKITELQLKLESALTELOL 593  
 QY 468 RKS-VPDCVLYLYLTKKENKALVRNYG-----KRGGRNQQLAR 507  
 DB 594 RKSQOHQMLVDSIVRQDRMYRILLSQTTGVAIPUHASLDDVSLASTPKRPSTQSTVST 653  
 QY 508 PSQEEKVPEEKEDKAE-----KTEKKKEKKDEEKDEKEDS-----KENTK 549  
 DB 654 PAPVPVISTEAIKAKALKQLQEIFENYKKEAENKIQNEOLEKLOEQVYTDLSQNTK 713  
 QY 550 EKDIDGTAEETE-----EREQATPRGRKTANSQGRKRGKGRITSMTNE 592  
 DB 714 ISTOLDFAKRYEMLQDNVEGYRREITSILHERNQ-----KLTTATQ--KQEQIINTMQD 766  
 QY 593 AAAASAAAAATEPPPLPPPPPEISTEPVETSWTEEMEVAKKGLVEHGRNWAATAK 652  
 DB 767 LRGAKEKLAVAEVR-----AENLKKEK-EMKLSEVRLSQOBSLLEABORGONLLT 817  
 QY 653 MVGT-----KSEAOCKNFYNY--KRRHNDLNLQHQKTSRKRPREEDVSCESVAS 704  
 DB 818 NLQTIQILSERSETKQRLSSQIEKLEHSHL-----KKLENEVEQRTILTR 867  
 QY 705 TVSAQEDIEDASNEENPESEVAVKPSDSPENATSR---GNTEPAVELEPTETAP 761  
 DB 868 NLDVQLDTRQLDPTNLHLTKELLNAQ--KEIATLKQHLSNNEVQVASSQSORTGK 925  
 QY 762 STSPSLAVPSTKPAEDSV-----ETOVNDSISAETAQMDVDQOEH--SAEESVCDP 813  
 DB 926 GO-----PSNKEDVDDLVSQRLQTEQVNDLKERLKTSTSNVEQYQAMVTSLESLNKE 979  
 QY 814 PPATKADSVDEVRVPEN-----HASKVSGDNTKERDLDRASEKVEPREED 859  
 DB 980 KQVTEEVKNIEVRLKESAEFTQLEKLMVEVEKEQELQDDKRAIESMEQQLSELAKT 1039  
 QY 860 LVVAQ---QINARPEPOSNDSSATCSADEVDGPEPERQRMFMDSKPSLLNPTGSIIV 916  
 DB 1040 LSSVQNEVQEQALQRASTALSNEQQAARRDCQEQAKTAVEAQNKYERE----- 1085  
 QY 917 SSPLKPNPLDLPOLQHRAAVTPPMVVSCTPCNIPITGTPVSGYALYQRIKAMHESALLEEQ 976  
 DB 1086 ---LMLHAADYEAQAAKEQVSKMAS-----VROH-----LEET 1116  
 QY 977 RORQEQIDLECRSSSTPCGTSKSPNREWE---VLQAPAPHOLI---TNLPGEVRLPTTPR 1029  
 DB 1117 TQKAEQLLEKAS-----WEERMLKDEVKVCRCEDLEKQNRLLHQ- 1162  
 QY 1030 TRPPPPPLIPSSKTTVASEKPSFIMGGSISQGTPTGTYLTSNQASYTQETPKPSVGSISLG 1089  
 DB 1163 -----IEKLSKVVASVKEGVQGPLNVSLSEEGKSQE---QILEILRF 1202  
 QY 1090 LPROESAKSATLPYIKOEESPSQNSQOPELLVRAQHEGVVVRGTAGIAQEGSTITRGP 1149  
 DB 1203 IRREKETAET-----RFEVAQVESLRYRQVVELLER---ELOE----- 1237  
 QY 1150 TSKISVESIPSLRGSITQGTPTALQGTPTALYKGSISRMPIEDS-SPEKREEAASKG 1208  
 DB 1238 -----LEDSLNAEREKQVQVTAKT 1255  
 QY 1209 HVIIYEGSGHILSYDNIKNAREGTRSPRTAHEISILKRSYESVEGNIKQGMNRPSPVAP 1268



Db 1256 MAQHE-ELMKKTETNNVYMETNK-MLREKERLEQDLQ-----MQAK 1296  
Qy 1269 LEGLLCRALPGSHDILKERTVLGSGTMOGTPRATTSTFEDGLKPKQIKRESPIRAF 1328  
Db 1297 VRKLELDLPLQEAANAELSEKSGMLQAKLLLEEDVKRWK-----1336  
Qy 1329 EGAIKGRPYDGIITIKEMGRSHEIPRO-DILTQESRK-----TPEVVQSTRPI-- 1377  
Db 1337-----ARNOHLVSOOKDPDTEYRKLLSEKVEVHKRLOOLTEELGR 1377  
Qy 1378 IEGSISQGTPIKFDNNSQSAIKHNVKSLITGPKSLRGMPPLEI-VPENIKVVERGK-- 1434  
Db 1378 LKAEIARSNASLTNNQNLQSLKEDLNKVRTEKETIQDKLDAIIDIOEKVKTITQVKKI 1437  
Qy 1435-----YEDVKAGETVRSRHTSVSSCP-----SVLRSTLHEAP----- 1467  
Db 1438 GRRYKTEYELKAQD-KVMETSAQSGDHQFQHVSVQEMQELKETLNOAETKSKLSQ 1496  
Qy 1468-----KAOLSPGIYDTSARRTPVSYONTMSR-GSPMMNRTSDVTIPPNKSNHKKSLT 1522  
Db 1497 VENLOKTLSEKETARNLEQTVQLQSELRLRQDLQRTTQEEQLRQIITEKEK----- 1552  
Qy 1523 PTQRESIPAKSPV---PGVDVSVSHSPDPHRRSTAG-----EYVWSHLPTQLDPAMP 1573  
Db 1553 -TRKAIVAAKSKIAHLAGVKDQLTRENEELKQNGALDQKDELDRITALKSQVEGRIS 1611  
Qy 1574 -FHRALDPAAYLAAYLFORQLSTPGYPSQYLYAMENTQTLINDYITQOQVNLRPDVA 1632  
Db 1612 RLRELREHQRHLQRD---EPOEPS-----NKVPQEQRIITLKTTPA 1652  
Qy 1633 RGLSPREOPLGLPYPATRGIIIDLTNMPPTILVPHPGGTSTPPMDRITVIGTQITFFPRP 1692  
Db 1653 SG-----ERGIASTSDPTANIKTPPVVSTVSKVTAAAMANKST--PRA 1695  
Qy 1693 YNSASMPG---HPTHLAAA---SAERERERERERERERERERERERERERERERER 1746  
Db 1696 SIRPMVTATVNTPTTATVMTTQVESQEMQSEGVEHVPVFGS-----1743  
Qy 1747 RPSGIVYRSPSVRTQETMLQRPVSFOGTNGTSVTP-----LDPTAQLRIMPLP 1799  
Db 1744 ---TSGSVRSTSPNV---QPSISQILTVQOQTQATFQVPTQOQHPQLEPANQ-----ELS 1794  
Qy 1800 AGGPSISQGLPASRYNTAADALAALVDA--AASAPQMDVSKTESKHEAA-----1847  
Db 1795 SNIVEVQSSPVERST-AVFGVVSATPSSSLPKRTREEEDSTIEASQVSDDTVEM 1853  
Qy 1848 RLEENLRSAVSAVSOQLEOKTLEVERKSVQCLYTSSAFPSGK-----POPHSSV 1899  
Db 1854 PLPKLKSVTPVGTBEEVMAEESTDGEVET-QVYNQDSQDSIGEGVGTQDYTPMEDSEET 1912  
Qy 1900 YSEAGKDKGPPPKSRYEELRTRGKTTITA---ANFTDVII-----TRQIASDK 1946  
Db 1913 SOSLOIDLGP-----LQSDQTTSSQDGKGKGDVIVIDSDDEEEDDDDD 1963  
Qy 1947 AREGQSQSSSSSLSS-----LQSDQTTSSQDGKGKGDVIVIDSDDEEEDDDDD 1972  
Db 1964 DTGMDGEDSNEGTSAGNDGNDGDAEGDGTDPCTETESMGGEKNHRAADSONS 2023  
Qy 1973 IEVISPASSAPPQKLTQYQYEVVVKANQANDPRTQVEGLPHHYRPOQESPSQOQLPP 2032  
Db 2024 GEGNTGAAESFSQVSEVREOQPSASERQAPRAP-QSPRRPPHPLPRLTTHAPPQELGP 2082  
Qy 2033 SSOABGMQVPRTHLRLTLADHICQILTDFAFNOVSSQTQQPTTFONSPSALVSTP 2092  
Db 2093 PVQRIQWTRRSVGRGLQLTGIGG-MQHFHFDDE-----DRTVPSTPLV--P 2129  
Qy 2093 VRTKTSNRYSPESQAQSVHHQPGSRVSPENLVKSRGSRPKSPERSHVSSEPIRISP 2152  
Db 2130 HRT-----DGAFAEAIH-----SPQ---VAGVPRFRFGP 2154  
Qy 2153 PQ-VPVVEHKKDLSLLLSQRA-----EPAEQRNDARSPGSIYLPSPFFTKLENTSPMK 2206  
Db 1255 PEDMPQTSSSHSLDGLQASQGLGMYETPLFLAHEEESGG-----RSVPTTPLQV 2204  
Qy 2207 SKKEIFRKLNSGGSDMAAQAQPGTEIFNLPVTTTS-GSVSSRGHS-----PAD 2256  
Db 2205 AAPVTVFTESTTS---DASEHASQ-----SVPVMTTSTGLTSTNETATGDDGDEVFE 2255  
Qy 2257 P-----ASNGLGDIIRKALMGSPDDKVEDGHVMSQPMGVVPGT-----ANTSVVTSGE 2306  
Db 2256 AESGISEAGLE-----IDSQEEEPVQASDES-LPSTSQDPPSSSSVDTs-- 2302  
Qy 2307 TRREEGDFSP 2316  
Db 2303 ----SSQPKP 2308  
RESULT 18  
BAT2\_HUMAN  
ID BAT2\_HUMAN STANDARD; PRT; 2142 AA.  
AC P48634;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2).  
GN BAT2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=T-cell;  
RX MEDLINE=90192810; PubMed=2156268;  
RA Banerji J., Sands J., Strominger J.L., Spies T.;  
RT "A gene pair from the human major histocompatibility complex encodes  
RT large proline-rich proteins with multiple repeated motifs and a  
RT single ubiquitin-like domain.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).  
RN [2]  
RP SEQUENCE OF 1-1860 FROM N.A.  
RX MEDLINE=93272029; PubMed=8499947;  
RA Iris F.J.M., Bougueret L., Prieur S., Caterina D., Primas G.,  
RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,  
RA Cohen D.;  
RT "Dense Alu clustering and a potential new member of the NF kappa B  
RT family within a 90 kilobase HLA class III segment.";  
RL Nat. Genet. 3:137-145(1993).  
CC -!- FUNCTION: UNKNOWN.  
CC -!- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
DR EMBL; M33509; AAA35585.1; -;  
DR EMBL; M33518; AAA35586.1; -;  
DR EMBL; M33512; AAA35586.1; JOINED.  
DR EMBL; Z15025; CAA78744.1; -;  
DR PIR; B35098; B35098.  
DR PIR; S36152; S36152.  
DR MIM; 142580; -;  
KW Repeat.  
FT DOMAIN 519 524 POLY-PRO.  
FT DOMAIN 636 657 GLN-RICH.  
FT DOMAIN 684 688 POLY-PRO.  
FT DOMAIN 699 704 POLY-PRO.  
FT DOMAIN 814 821 POLY-PRO.  
FT DOMAIN 1340 1345 POLY-GLY.  
FT DOMAIN 1398 1403 POLY-GLY.  
FT DOMAIN 1436 1442 POLY-PRO.



FT DOMAIN 1982 1991 POLY-PRO.  
FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.  
FT REPEAT 41 95 1-1.  
FT REPEAT 98 154 1-2.  
FT REPEAT 281 337 1-3.  
FT REPEAT 1740 1795 1-4.  
FT DOMAIN 337 549 2 X TYPE B REPEATS.  
FT REPEAT 337 418 2-1.  
FT REPEAT 476 549 2-2.  
FT DOMAIN 1899 2089 3 X 50 AA TYPE C REPEATS.  
FT REPEAT 1899 1948 3-1.  
FT REPEAT 1965 2014 3-2.  
FT REPEAT 2040 2089 3-3.  
FT CONFLICT 57 57 R -> A (IN REF. 2).  
FT CONFLICT 109 109 Q -> S (IN REF. 2).  
FT CONFLICT 414 414 P -> PPHRGAGNMGPP (IN REF. 2).  
FT CONFLICT 532 532 T -> K (IN REF. 2).  
FT CONFLICT 682 682 Q -> K (IN REF. 2).  
FT CONFLICT 730 730 E -> D (IN REF. 2).  
FT CONFLICT 750 750 L -> R (IN REF. 2).  
FT CONFLICT 834 834 A -> T (IN REF. 2).  
FT CONFLICT 1035 1035 G -> A (IN REF. 2).  
FT CONFLICT 1068 1068 M -> L (IN REF. 2).  
FT CONFLICT 1285 1285 P -> R (IN REF. 2).  
FT CONFLICT 1400 1400 T -> S (IN REF. 2).  
FT CONFLICT 1611 1611 G -> A (IN REF. 2).  
FT CONFLICT 1729 1729 G -> A (IN REF. 2).  
SQ SEQUENCE 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;

Query Match 2.8%; Score 353; DB 1; Length 2142;  
Best Local Similarity 20.4%; Pred. No. 3.8e-06;  
Matches 439; Conservative 224; Mismatches 833; Indels 656; Gaps 104;

Qy 503 QZIARPS---QEKVEEKEEDAEKTEKEEKEEKEDEK-----EKEDSKENTKEKDKI 554  
Db 299 EPVGRPSILKEDNL-----KEFDOLDQENDGAGAAEVDYTEKLFKFSDEE 345

Qy 555 DGTAEETE-----EREQATPRGRKNTANSQGRKRITRS--MTNEAAAAAASAAAAATE--- 605  
Db 346 DGRSDDEGAEGHRDSQSASGEERPPADGKGNSEPPPTKTAWAETSRPTEPEGP 405

Qy 606 -EPPPPPLPP-----PPPISTPEVETSRWTEEEEAQKGLVHGRNWAATAK 652  
Db 406 PAPKPPPLPGDYDPRGGPPCKPAPADEDEAWQRKQSSSEISL--AVERAR----- 456

Qy 653 MVGTKEAQCNFNFTYKRRNLNLLQHQKTSRKPREDRDVSQCESVASTVSAQDE 712  
Db 457 -----RRREERMRQERRAACAEKL-----KRLDE 483

Qy 713 DIEASNEENPEDSEVAVKPSEDSPENATSR---GNTEPAVELEPTTETAPSTSPSLAV 769  
Db 484 KFGAPDKLKAEPAPPA--PSTAPPAPVAPKELPAPPAPPPASAPTEPEPE--EPAQAP 541

Qy 770 P--STKPAEDESVEQVN-----DSISAEATAQMDVDQOEHSAE--EGSVCDPPPPAT--- 817  
Db 542 PAQSTPTPGVAAAPTLYSGGGSSTSSGSPSEASPVPEQLPSKEGPEPEVPPPTTPV 601

Qy 818 -----KADSDVDFVRVPHENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQQAQNAQPE 872  
Db 602 PKVEPKGIGFTQPP-----SGLGYPKYQKSL---PPRFQRQQEQQLLQKQ---QQHQ 651

Qy 873 PQSDNSDTSACADEVDGEPEQRMFMDSKPSLLNPTGSGILYSSPL--KENPLDLPOLQ 931  
Db 652 WQHQOQSAPPTPPPPQPVQTVLGAVPAQAPP---PPPKALYPGALGRPPPPPMNFD 708

Qy 932 HRAAVIPPMWCTPCNIPGTVSGYALYQRHIKAMHESALLEBQKQ-----ROEQIDLE 986  
Db 709 PRMMIPPYVD--PRLLQGRPPLEFP-----PGVHPSGLVPRERSDSLGSSEPFDRH 760

Qy 987 CRSSTSPCGTSK--SPNREW-----EVLQAP-----HQLITNLPDGVRLPPTTRPRPP 1033  
Db 761 APAMLRERGTTPVPDKLAWGVDFVETATPAEPRLTSLPLRQAADDDKGM--SETPPVPPPP 819

Qy 1034 PPLIPSSKTTVASEKPSFIMGSSISQGTPTGYLTASHNQASYTQETPKPSVGSISLGLPRQ 1093  
Db 820 PPYLAS-----YGFEPENGAGPPI---SRFPLEECPRLPWP-----PGS 858

Qy 1094 QESAKSATLYIKOEFPSPRSQSOPE--GLLVRAQHEGV-----VRGTAGIAQSGS 1143  
Db 859 DEVAKIQTTPPKPEKPEETAQLTGAAGRLPASRSAGAGPPPPRRRESRTETRMGPRGS 918

Qy 1144 ITRGPTSKISVESIPSLRGSITQGTALPQGTPTAEALVKGSISSRMPIEDSSPKGREE 1203  
Db 919 SRGGIPPEEGAP--PRRAGPIKKPPPTKVEELPPKPLEQG-----DETAKPPAPD 968

Qy 1204 AASKGHVIEGKSGHILSYDNIKNARECTSPRTAHEISLKRYSVESGNIKQKMSMRES 1263  
Db 969 PLK---ITKGLG-----GKPEPPNGNLSP--APRLRDYSYE-----RVG 1005

Qy 1264 PVSAPLEGLICRALPRGSPHSDLKERTVLSGIMOGT-----PRATESFEDGLK 1313  
Db 1006 PTS-----CRGRGRGEYFA-----RGRGFRGTGGRGRGQQAASAVTESFEEM-- 1049

Qy 1314 YPKQIKRESPIRAFEAGITKGYDGIITIKEMGRSITHEIPRODILTQESRKTPEVQOS 1373  
Db 1050 -----GV-----EVG-----OGDOTLLLLPEA 1066

Qy 1374 TRPIIEGSIQGTPIKFDNNSGQSAIKHNKSLITGPSKLSRGMPPLEIVPENIKVVERG 1433  
Db 1067 AMPARHGARGVSMR--KSPSGAGSQAQQAARPMRVIWLLQTRLP----- 1110

Qy 1434 KYEDVKAQETVRSRHTSVVSSGSPVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMS 1493  
Db 1111 -----HPRREHSRSPSRSTPRTLHRAPARFTCPGVGESSLP-----EGAIS 1154

Qy 1494 RGSPMMNRTSDVT---IPPKNSTNHERKST--LTPQRESIIPAKSPVPGVDPVSHSPDP 1549  
Db 1155 PGRRRREAPPQVCPGWSPPAKSLAPKPTGTLPPSKPLKEKL--IPG-----PLSP 1205

Qy 1550 HHRGTAGEVWSHLPTQOLDPAMPFHRLDPAALAAAYLQRLQSLPTPGVPSQVLYAMENT 1609  
Db 1206 VARGSGNG---GSNVGME--DGERPFRRHGRA-----QQQ-----DKPPFRRLKQE-- 1248

Qy 1610 ROTILNDYITSQOMQVNLRPDVARGLSPREQPLGLPYPATRGIIDLTMNMPITILVPHPG 1669  
Db 1249 -----RENAARG--SEKPKSLTLPASA-----PG-- 1270

Qy 1670 TSTPPMDRITVIGTQITFPPRPYNASAKSP-----GHPTHAAAAAASERERER 1720  
Db 1271 -----PEEAL-----TTVTAPAPRAAAKSPDLNSQNSDOANEEMETASESDFTSERRG 1321

Qy 1721 EKE-----RERERIAAASDLYLRPGSQPGPSHG 1752  
Db 1322 DKEAPPPVLLTPKAVGTPGGGGGAVPGIISAMSRGDLISQRAKDLKSRFSQ--RPGMER 1379

Qy 1753 YVRSPSPSVRTQETMLQORPSVFQGTNGTSV-----ITPLDPTAQLRIMPLPAGGSPISQ 1808  
Db 1380 QNRRPQPGGKA-----GSSGSSGGGGGGPGGTG-----PGRGDKRSM 1419

Qy 1809 LPASRYNTAADALAAALVDAASAFQMDYSKTESKHEARLEENLRSAVSEQQOLEQ 1868  
Db 1420 SPKNRSRPPEE-----RPPGLPLPPPPSSSAVFLDQVTHSNPAGI--QALAAQ 1467

Qy 1869 KTLVEKRSVQCLYTSSAFFSGKPPQHSSVVYSEAGDKGPPKRSRYEELTRGKTTIT 1928  
Db 1468 ---LSSRQGSVT---APGCHPRHKFGP-----POAQGPSRPRTTRYEPQVNSG----- 1511

Qy 1929 AANFDVITITQIASDKDARERGSSDSSSLSSHRYETPSDAIEVISPASSPAPPOEK 1988  
Db 1512 -----LSSDPHFEPGPMVRGCG-----TPRDSAGV-----SPFPKRR 1546

Qy 1989 LQ--TYQPEVVKANAENDPTROYEGPLHHYRPOQESPPSQOOLPPSSQAEGMGVPTHR 2047  
Db 1547 ERPPKPELL---QESLPPPHSHSGFL---GSKPEGPGQAE-----SRDTG 1587



QY 2048 LITLADHICQIITQDFARN--QVSSOTPOQPTSTFQ-----NSPSALV 2089  
Db 1588 TEALTPIHNLHTATSRKSYRPTSMEPWEPLSPFEDVAGTQMSQSDSGVDLSGDSQVS 1647  
QY 2090 STPVTKTSNRYSPESQASVHH--QRPGRSVSPENLV--DKSRGSRPGKSP-----2137  
Db 1648 SGP-----CSQSSSPDGLKGAAGPPKPRPGGS--SPLNAVPCGPPGSEPPRRPPAPHDG 1702  
QY 2138 ERSVSSPEYPIPPQVPVHVKODSLLLSQSGAPAEORNDARSPGSIYLPSEFTK 2197  
Db 1703 DRKELPRE--QPLPFG--PCTERSQR-----TDRTGTEPGPIRPSHR--PGP-----PVQFGT 1749  
QY 2198 LENTSPM-----VSKQKQIFRKLNSGGGSDMAAOPGTIFN--LPATVTTSGSVSS 2249  
Db 1750 SDKSDRLVVGDSILKAEKE-----LTAS-----VTEAIPVSRDWELLPSAASAEPOS 1798  
QY 2250 R-----GHISFADPASN--LGLEDIIRKALMGSDFKVDKEDHGVVM-----2286  
Db 1799 KNLDSGHCVPEPSSSGORLYPEVFGSAGPSSQISGSGHLSITSKOMRLRPGTSLHP 1858  
QY 2287 --SOPMGVVPCTANTSVVTSG-----ETREEGDPSPHSGGVCKPKLISKNSRK 2334  
Db 1859 YRSQPLVLPGPAPPSALLSGVALKQFLDFSTWQATLGLPAGGVLYPPSPFLYSPAF 1918  
QY 2335 SKSPIGOGYLGTERPSVSVSHSEGDYHRO--TPGNWADERPSTGSTOF-----2383  
Db 1919 CPSPLDTSLLQVRQ-----DLPSDFSYSTPLOGPGSGFLPSGAPAAQOMLLPMVDSQL 1973  
QY 2384 -----PVNPLTMRMLSTPTTACAPSAYNQAPHOQNIWEREAP 2426  
Db 1974 PVNFGSLPPAPPAPPLSLPLVGPALQPPESLAVRPPAPATRVL-----PSP 2022

RESULT 19  
YDBJ\_SCHPO  
ID YDBJ\_SCHPO STANDARD; PRT; 661 AA.  
AC Q10369; Q10226;  
DT 01-JUL-1996 (Rel. 34, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 75.1 KDA PROTEIN C22E12.19 IN CHROMOSOME I.  
GN SPAC22E12.19 OR SPAC2E12.01.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE OF 1-230 FROM N.A.  
RC STRAIN=972;  
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 215-661 FROM N.A.  
RC STRAIN=972;  
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 215-661 FROM N.A.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Z70043; CAA93906.1; -  
DR EMBL; Z69726; CAA93545.1; -  
DR InterPro; IPR001005; -  
DR Pfam; PF00249; myb\_DNA-binding; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 661 AA; 75096 MW; 59A15D8C990E2C45 CRC64;

Query Match 2.8%; Score 350; DB 1; Length 661;  
Best Local Similarity 21.8%; Pred. No. 1.3e-06;  
Matches 155; Conservative 122; Mismatches 298; Indels 136; Gaps 28;  
QY 224 SPPPV-----EQKHSIVQ-----IYDENRKAEEAHKIFEGIGPKVPELPL 266  
Db 28 SEPVGMDNEDSDTNTSITQSPNSEKLTDLQESDTRALO--EKLQNIYALTQOLF 86  
QY 267 NQPSDTKYHENTKTNVMRKLLFFKRNHARKQREKICQYDQIMEAWKVKVDRIE 326  
Db 87 KNVEDSYFNLRKFKORDQTIIVGVNKRHRHVLNKKIKRLQSHWKQVLRWEENIARVD 146  
QY 327 N-NPRRAKESKTREYKQFPEIRKQREQOQRFQVQRGAGLSATARSEHEISLID 385  
Db 147 RLTEIDTKNAKKESEPIK-----RSTRKVMSNF-----TAGDIVRSEEFLELA 192  
QY 386 GLSQE---NNEKOMQLSIPPMFMFAEQRRVKFIN--MNGLMEDPMKYKDRQFMVWT 441  
Db 193 KLEQKEASNVSEASRIATIPPMILSEEEVKSOYFNDQSLVTDCKPKYHFQSMPDIN 252  
QY 442 DHEKEIFKDFPIQHPKNGFLIASYLERKSVDPDCVLYIYLLKKNNYKALVRNRTKGRGR 501  
Db 253 EEOHSIFVQOQFILHGKFGKIAEAVPGKSKCEVLHYLTAKRTDYRALV--ASATKTGR 311  
QY 502 NQIARPSQE--EKVEEKEEDKAETKEKEKDEEKEDEKSKENTKEKDKIDGTAE 560  
Db 312 RRKLLPSQGGKKKSGSALWVDIEAADINKTEENINNOFOEASVTADNMNTWDNT--PS 370  
QY 561 TEEREQATPR--GRKTANSQGRRKGRITRSMTNEAAAAAASAAAAATEPPPLPPPPPEPIS 619  
Db 371 VENVESANENVNHNADQMDK---IKSLV--EGNSAYEIEKGAQ-----PDPM 417  
QY 620 ---TEPVETSRWTEEMEVAKKGLVEHGRNWAATAKAVGKTKSEAOCKNFYN---YKRRH 673  
Db 418 IDMTDKSETVSGFKHDVVDYTAENEGNNTLLQIKESVHEKTPQDEPMDISQDTIKQED 477  
QY 674 NLDNLLQOH--KOKTSRKPREERD---VSQESVASTVSAOEDIEASNEENPEDSE 727  
Db 478 YYEPKLEQHSKKNSISTRKEADAALANLSAVGRSISA-----VDESA 523  
QY 728 VEAVKPSDESPENA-----TSRGNTPEPAVELEP--TTETAPSTSPSLAV-----PSTKPAED 777  
Db 524 HGHLPQWDEKEEALIFSLAQGMNPKMPLTPRRASTGPRPRTFQLTEIDSPNRRASD 583  
QY 778 -----ESVETQVNDISISAETAQMDVQOQHSAAEESVCDPPATKADSVDEVVRYPEN 831  
Db 584 CITPSISKILKMVSEDASKSRIDELSVEDEQHT-----T 617  
QY 832 HASKVEGDNT---KERDLDRAS--EKVEPRDEDLVVAQOINAAQRPESQSDND 878  
Db 618 HSSHTTSDINAFPSQSFPFRASIHITLAALGEDIV-----ERQSKND 658

RESULT 20  
YKZ6\_CAEEL  
ID YKZ6\_CAEEL STANDARD; PRT; 1018 AA.  
AC P34333;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE HYPOTHETICAL 113.7 KDA PROTEIN C14B9.6 IN CHROMOSOME III.  
GN C14B9.6  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,



























RA Kindler S., Schwanke B., Schulz B., Garner C.C.;  
RT "Complete cDNA sequence encoding rat high and low molecular weight  
MAP2.";  
RL Nucleic Acids Res. 18:2822-2822(1990).  
RN [2]  
RP SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.  
RC STRAIN-WISTAR; TISSUE=Brain;  
RX MEDLINE=91060576; PubMed=2174050;  
RT Kindler S., Schulz B., Goedert M., Garner C.C.;  
RA "Molecular structure of microtubule-associated protein 2b and 2c from  
rat brain.";  
RL J. Biol. Chem. 265:19679-19684(1990).  
RN [3]  
RP SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.  
RX MEDLINE=90221819; PubMed=2326166;  
RT Doll T., Papandrikopoulou A., Matus A.;  
RA "Nucleotide and amino acid sequences of embryonic rat MAP2c.";  
RL Nucleic Acids Res. 18:361-361(1990).  
RN [4]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=89365159; PubMed=2770869;  
RA Papandrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;  
RT "Embryonic MAP2 lacks the cross-linking sidearm sequences and  
dendritic targeting signal of adult MAP2.";  
RL Nature 340:650-652(1989).  
RN [5]  
RP SEQUENCE OF 1695-1725 FROM N.A.  
RX MEDLINE=94110302; PubMed=8282767;  
RA Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;  
RT "An isoform of microtubule-associated protein 2 (MAP2) containing  
four repeats of the tubulin-binding motif.";  
RL J. Cell Sci. 106:633-640(1993).  
CC -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY  
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO  
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.  
CC -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY  
CC ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR  
CC FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.  
CC -!- DEVELOPMENTAL LAGS: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN  
CC DEVELOPMENT AND UNTIL POSTNATAL DAY 10. MAP2B IS EXPRESSED  
CC THROUGHOUT BRAIN DEVELOPMENT.  
CC -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X51842; CAA36135.1; -;  
DR EMBL; X17682; CAA35667.1; -;  
DR EMBL; X71487; CAA50588.1; -;  
DR PIR; S07887; S07887.  
DR PIR; S10003; S10003.  
DR PIR; A37981; A37981.  
DR InterPro; IPR001084; -;  
DR Pfam; PF00418; tubulin-binding; 4.  
DR PROSITE; PS00229; TAU\_MAP\_1; 3.  
KW Microtubules; Repeat; Alternative splicing; Calmodulin-binding.  
FT DOMAIN 1454 1474  
FT REPEAT 1664 1694 TAU/MAP MOTIF.  
FT REPEAT 1695 1725 TAU/MAP MOTIF.  
FT REPEAT 1726 1756 TAU/MAP MOTIF.  
FT REPEAT 1757 1788 TAU/MAP MOTIF.  
FT VARSPLIC 152 1514 MISSING (IN ISOFORM MAP2C).  
FT VARSPLIC 1695 1725 MISSING (IN ISOFORM WITH 3 TAU/MAP  
SEQUENCE 1861 AA; 202409 MW; 42DCFF16D21EF54E CRC64;

Query Match

2.6%; Score 332.5; DB 1; Length 1861;

Best Local Similarity 20.0%; Pred. No. 1.9e-05;  
Matches 357; Conservative 237; Mismatches 588; Indels 603; Gaps 94;  
QY 1008 OPAPHQLITNLPPEGVRLPTTRTPPPPLIPSSKKTVAASEKSFIMGSGISOGTPTCTYLT 1067  
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 119 OPAALPLAAE--ETVNLPP-----PSPPPSPASEQTAALAE-----LL 154  
QY 1068 SHNQASYTQETPKPSVGSISLGLPRQESAKSATLPYIKQEEFSPRSONSOPEGLLVRAO 1127  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 155 TASKMEFFPEQQLPS--SFAEPLDKEETEFKMSKP---GEDFEHAALVPQD----- 202  
QY 1128 HGVVVRGTAGAIQ-----EGSITRGTPTSKISVESIPSRGSIQTGTPALPOTGIPT 1179  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 203 -----TSKTPQDKKQDMEGEKSPASPPFAQTGCTNLEDIK-QITEPSITVPSIGLSA 254  
QY 1180 EALV---KGSISRMPIDSSPEKGREEAASKGHYI-----YECK----- 1215  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 255 EPLAPKQDKDMFIEMPVESKDEWGLAAPISPGPLTPMREKDVLEIDIPWEKQKQFDSMP 314  
QY 1216 -----SGHILSYDNIKNAK--EGTRS-----PRTAHEISLKRYSVESVEGNI 1254  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 315 SPFHGGSFTLPDLYKDERVTEGSPFAPVFQSDDKMSLQDTSGSATSKESKDEEPK 374  
QY 1255 KQGMSPRESVPASPLEGLICRALPRGSPHSOLKERTVLGSGSIMQGTTPRATTFEDGLK 1314  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 375 DKADKVADVVPVSEATTVL-----GDVHSPAPEGFV--GENISGEKCTTD----- 417  
QY 1315 PKQIKRESPPRAFEGATKPKPYGITTIKEMGR-SHEIPRODILITQESKRTPE----- 1369  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 418 --QEKETSTPSVQEPITLTETEPQ---TKLEETSKVSEETVAKE---BESLKDDKAG 469  
QY 1370 VVQSRPIIEGSGISOGTPIKFNNSGO---SAIKH---NVKSLITGPSKSLRGMPL 1420  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 470 VIQTS---TEQFSK-----EDQKQEQTIHALKQDSPISLEQAVTDAAMATKLEK 520  
QY 1421 EIVPENI--KVVERGKEDVKAGETVRSRHTSVSGSPSVLSTLHEAPKAOLSPGI--Y 1476  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 521 TSEPEAVSEKREIOGLFEEDIAK-----SKLEGAGSATVAEV--EMPFYEDKSGMSKY 572  
QY 1477 DTSARRTPVS-----YQNTMSRSGSPMMNRISDVITPNKSTNHERKSTLPTQRE 1527  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 573 FETSAKEDVTRSTGLGSDYIELSDRGN--AQESLDTVSPKNQ---QDEKELLAKASOP 627  
QY 1528 STPAKSPVGVDPVVSHPDPDHRG-STAGEVYVNSHLPTQL--DPAMPFHR--ALDPAA 1582  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 628 SPPA-----HEAGYSTLAGSYSDHPSELPEEPSPQERMFIDPKV 669  
QY 1583 AAYLFQRLSPTPGYPQVQLYAMENTRQITILNDY-----ITSQQMVNL-----RPDV 1631  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 670 -----YGEKRDLSKKNKDDLTLSRLSLGLGGRSAIEGRSMINLPMSCLDLSI 715  
QY 1632 ARG-----LSPREQPLGLPYPATRGIIDLTN--MPPTILVPHPGGTSTPMDRITVI 1681  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 716 ALGFNFGRHDLSPLASDI---LTNTSGMDGDDYLPPT-----TPAVEKIPCF 762  
QY 1682 P-----GTQIT-----PPRPY--NSASMSGHPHTLAAA----- 1709  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 763 PIESKEEDKTEQAKVTGGQITQVETSSPPPAKEYKNGTVMAPDLPMLDLAGTRSR 822  
QY 1710 -ASABER 1755  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 823 LASVSADAEAARVARKSVSEAVVAESS--TGLPPVADDSQPVKPDQLEDMGVCVFNKYTP 881  
QY 1756 SPSPSVRTQETMLQORPSVFGTNGTSTVITPLDPTAQRLIMPLPAGGSGISOGLPASRYN 1815  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 882 LPSP-VQDSNLSGSGSYFEGTQD-DKYRRDLATDLSLIEVKLAAAG----- 926  
QY 1816 TAADALAALVDA--ASAPQMDVSKTKESKHEA-----ARLEENLRSR-----SAAV 1860  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :  
Db 927 RVKDEFTAEKASPPSSADKSGLSREFQDQRKANDKLDLVLEKSEHVDKSHAKESSEV 986  
QY 1861 SEQQOLEQKTLVEKRSVQCLYTSSAFPSGKPPQPHSSVVVYSEAGDKGPPPKSYEEELR 1920  
Db : : : : : || : : : : : || : : : : : || : : : : : || : : : : :



Db 987 GDKVFLGLVGYTEQTSAKELITT-----KETAP-----ERAEGKLS 1023  
QY 1921 T-----RCKTITITANFIDVITITROIASDKDAR--ERGSQSS-----DS 1957  
Db 1024 SVPEVAEYETTKADQGLDVAARKDDQSPDKIVSDFGOMASGMSVDAGKTIKLFVDQ 1083  
QY 1958 SSSLSHRYETPSDAIEVISPASS-----PAPPOEKLOTQYQPVVKANAENDPTROY 2010  
Db 1084 QLTLS-----EAPQETDSEWGTIESHVKGAKVSETEVEKVKAPDLV--HQEAVDKESY 1138  
QY 2011 EGPLHHYRQOES-----PSPQOOL-----PPSQAEGMGQVPR 2045  
Db 1139 ESSGESHLSITKDPKGGKETSPTSLSIQDEVALKLSVEIPCPVPVSEAD----- 1190  
QY 2046 HRLITLADHICQIITODAR--NQVSSOTPOOP--PTSTFQNSALYSTVPTVKTSNRY 2101  
Db 1191 ----SSIDEKAE--VOMEFIQLPKEESTETPDIPALPSDVTQPOQPAVVSEPAEVRGEE- 1244  
QY 2102 SPESAQAQSVHQR-----PGSRVSPENLVDKGRGSRPG----- 2134  
Db 1245 --EIEAEGEYDKLLFRSDTLQITDLLVPGSR--EEFVETCGEKGHVSVVITIEDFI 1299  
QY 2135 ----KSPERSHVSPEYE--PISPPQV--PVVHEKQDLSLLLSQRGAEPAEQNRDARSP 2185  
Db 1300 TVVQTDTDEGELGSHSVRFAAPVQPEERRPYPHDEEVLMAAQAQAEKPKDGDSPADAT 1359  
QY 2186 GSISYLP-----PMSSTPTPIACAPS--AVNQAAPHQON 2417  
Db 1360 PEKEVPSEYKTYDYDKETTTIDDSIMDADSLWVDTQDDRSILTEQLETIPKEERA 1419  
QY 2204 -----MVKSSKOEIPKLNLSGGG-----DSDMAAAGPOT-----EIFNLPAVTTSG 2245  
Db 1420 EKEARRPSLEKHKKEPKF-----TGRGRISTPERREVAKEPSTVSRDEVRKKAAYKKA 1475  
QY 2246 SVSS-----RGHSFADPASNLGLEDII-----RKALMGSD-----D 2277  
Db 1476 ELAKESEVQAHS-----PSRKLKPAIKYTRPHTLSCVRKKTATSGESAQAPSAFKQAKD 1532  
QY 2278 KVEDHGVVMS-----QPMGVVPGTANTSVTSGETTRREEDPSHSGGVCKPKLISK 2329  
Db 1533 KSTD-GIKYSEKRSLSRPSILPRRG-----VSGD--RENSFSLNS--ISS 1578  
QY 2330 SNRSKSPFPGQYGLTERPSSVSSV-----HSEGDYHROTPGNWAWEDRPSSTGTFPPY 2385  
Db 1579 ARRTTRSEPIRRAGSGTSTPTPGSTAITPGTSPYSYSSRTPG-----TPGPPSPYR 1630  
QY 2386 NPLTNR-----MLSTPTPIACAPS--AVNQAAPHQON 2417  
Db 1631 TPGTPKGLVPSKVAIIRTPPKSPATPKQLRLINQPLDLKN 1675  
  
RESULT 25  
AKAC\_HUMAN STANDARD; PRT; 1781 AA.  
AC Q02952; Q99970; Q00498; Q00310;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP  
250) (MYASTHENIA GRAVIS AUTOANTIGEN GRAVIN).  
GN AKAP12 OR AKAP250.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP TISSUE=Heart;  
RX MEDLINE=97153077; PubMed=9000000;  
RA Nauert J.B., Klauk T.M., Langeberg L.K., Scott J.D.;  
RT "Gravin, an autoantigen recognized by serum from myasthenia gravis  
patients, is a kinase scaffold protein.";  
RL Curr. Biol. 7:52-62(1997).

[2]  
SEQUENCE FROM N.A. (ISOFORM 2).  
TISSUE=Umbilical vein endothelial cells;  
MEDLINE=98269042; Pubmed=9604001;  
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;  
RT "Changes of gene expression by lysophosphatidylcholine in vascular  
endothelial cells: 12 up-regulated distinct genes including 5 cell  
growth-related, 3 thrombosis-related, and 4 others.";  
RL J. Biochem. 123:1119-1126(1998).  
[3]  
SEQUENCE OF 43-1781 FROM N.A.  
TISSUE=Umbilical vein endothelial cells;  
Bowditch R.D., Ginsberg M.H.;  
RT "Sequence of gravin cDNA isolated from a human umbilical vein  
endothelial cell library.";  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE OF 1477-1781 FROM N.A.  
TISSUE=Umbilical vein endothelial cells;  
MEDLINE=92395179; Pubmed=152245;  
Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,  
Lindstrom J., Ginsberg M.H.;  
RT "Molecular cloning and preliminary characterization of a novel  
cytoplasmic antigen recognized by myasthenia gravis sera.";  
RL J. Clin. Invest. 90:992-999(1992).  
CC -1- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR  
COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C  
(PKC).  
CC -1- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL  
CYTOSKELETON.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE  
PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED  
FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,  
MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.  
CC -1- INDUCTION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).  
CC -1- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE  
INVOLVED IN BINDING PKC.  
CC -1- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY  
PATIENTS WITH MYASTHENIA GRAVIS (MG).  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U81607; AAC51366.1; -;  
EMBL; AF001504; AAB58938.1; -;  
EMBL; AB003476; BAAL1927.1; -;  
EMBL; M96322; AAA35931.1; -;  
MIM; 604698; -;  
PIR; A43922; A43922.  
InterPro: IPR001573; -;  
InterPro: IPR001899; -;  
Antigen; Alternative splicing. AKAP 1.  
DOMAIN 603 633 AKAP 1.  
DOMAIN 752 782 AKAP 2.  
DOMAIN 797 827 AKAP 3.  
DOMAIN 98 101 POLY-GLU.  
DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).  
DOMAIN 1540 1553 MISSING (IN ISOFORM 2).  
VARSPIC 1 98 RII-BINDING (PROBABLE).  
VARSPIC 99 106 EEEVIVTE -> MLGITIT (IN ISOFORM 2).  
CONFLICT 117 117 E -> K (IN REF. 2).  
CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).  
CONFLICT 215 215 Q -> K (IN REF. 2).  
CONFLICT 448 448 G -> E (IN REF. 2 AND 3).  
CONFLICT 694 694 R -> G (IN REF. 2 AND 3).  
CONFLICT 867 867 G -> S (IN REF. 2 AND 3).















```
Qy 1005 -----EVLPAPHOLITNPEGVRLPTTPT 1030
Db 1566 GNFCPLDKCDDDDYESKMMQCKDRWHVHSCENLSDMEYELSNLPESVAVTCVNC 1625
Qy 1031 RPPPLIPSSKTTVAERKPSFIMGSGISQGTGYLTLSHNAQSYQTPKPSVSGSISLGL 1090
Db 1626 ERHP-----AEWRLALEKELQISLKQVLTALLNSRTTSH-LLRYRQAAPPDNLN----- 1673
Qy 1091 PROESAKSAILP-----YKQEFSP----- 1112
Db 1674 PETEESIPRSPEGDPPIVTEVSKQDDQDPLDLEGVKRMDQGNYSVLEFSDDIYKI 1733
Qy 1113 -----RSONSOP-----GLLVRA-----OHEGV-----VRGTAGAI----- 1139
Db 1734 IQAAINSGGPEIKKANSWMKSFIRQMERVFWFVSKSRFEPNPKVSSNSGMLPNVAV 1793
Qy 1140 -----QEGSITRGTP-TSKISVESIPSLURGSITQGTALPQTGIPTEALV 1183
Db 1794 LPPSLDHNVAQWOEREENSHTQEPPLMKKIIIPAPKPKGPEPDSPTPLHP-----PTPPIL 1849
Qy 1184 KGSISRMPIEDSSPKGPEE-----AASKGHVIEYK-----SGH 1218
Db 1850 STDRSREDSPENPPGIEDNRQCALCUTYGDSDANDAGRLLYIGQNEWTHVNCALWSAE 1909
Qy 1219 ILSYD--NIKNAR----- 1229
Db 1910 VFEDDDGSLKNVHMAVIRGKQRCFCQKPGATVGCCLTCTSNYHFMCSRAKNCFVLD 1969
Qy 1230 -----EGTRSPRTAHE-----ISLKRYSVEGNITKQGMRESVPS 1266
Db 1970 KKVYQQRHDLIKGEVWENGFEVRRVFDVDFEGISLRKF--LNGLEPENIHMMIGSMT 2027
Qy 1267 APLEGLI-----C--RALPRGSPHS-----DLKERVLSGSIMQGT-----RATT 1305
Db 2028 IDCLGILNDSCEDKLPYIGQCSRVVWSTDAKKRCVYTCCKIVECPRPVPEPDINSTV 2087
Qy 1306 E-----SPEDGLKYPKQIKRE--SPPIRAFGAITKGPYDGIITIKEMGRSI 1351
Db 2088 EHENDRTIAHPTSTFESSKESQNTAIIISP-----SPDRPHSQTSFGSYYHVI 2139
Qy 1352 HEIPR-----QDILQES-----RKTPEVOSTRPIIEGSIQGTPTKFDNNSQSIAI-----KH 1401
Db 2140 SKVPRIPTPSYPTQSPGCRPLPSAGSPT-PTTHIEVTVGDPLL-----SSGLRSIGSRH 2195
Qy 1402 NVKSLITGPSKLSRGMPLEIVPENIKVVERGYEDVKAGETV-RSRTSVVSSGPSV-L 1459
Db 2196 STSLSPORSKL-RIMSPN-----RTGNTYSRNNVSSYSTTGTAIDL 2236
Qy 1460 RSTL-----HEAPKAQLSPGIYDDTSA-----RRTPVSYQNTMSR-----GSPMMRTSDVT 1506
Db 2237 ESSAKVVDHVLGPLNSSLQNTSTSSNLQRTVTVGNKNSHLDGSSSEMKOSSASDL 2296
Qy 1507 IPPNKNSTHNRKSLITPTQRESIPAKSPVPGVDVWVSHSPDPHHRGSTAGEVWHLPT 1566
Db 2297 VSKSSSLGKERTKVLSSKSSSGSAHNAVPIPKLA-----PQVHNTTSRELNVSKIGS 2350
Qy 1567 QLDPAMPFHRALDPAAYALFOR-----QLSPTPGVPSQ-----YQLYAMENT 1609
Db 2351 FAEPSSVSFSKEALSFFHLHLQRNDRODHTDSTQANSPPDETEVKTLKLSGMSN- 2409
Qy 1610 ROTILNDYITSQOM-----QVNLRPDVA----- 1632
Db 2410 RSSIINEHMGSSRRDRQKKGKCKETFKKHSSKSFLEPGQVTTGEBGNLKPFEDEV 2469
Qy 1633 -----RGLS-----PREQPL-----GLPYPATRGI----- 1652
Db 2470 TPEYMGQPCNNVSDKIDGDLGSLMPGVKAPPQMGVEGSAKELQAPRKRTVKVTLPLKM 2529
Qy 1653 -----IDLTNNPPT-----TLVPHPGTS----- 1671
Db 2530 ENESQSKNALKESSPASPLQIESTSPTEPISASENPBGCPVAQSPNNTSCQDSQSNYQ 2589
```

## RESULT 27

```
PLEI_RAT
ID PLEI_RAT STANDARD; PRT; 4687 AA.
AC P30427; 008879; 008880; 008881;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLECTIN 1.
GN PLECT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
```



OX . NCBI\_TaxID-10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-91268156; PubMed-2050743;  
 RA Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,  
 RA Hauptmann R., Stratowa C., Stewart M.;  
 RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide  
 chain with a three-domain structure based on a central alpha-helical  
 coiled coil.";  
 RL J. Cell Biol. 114:83-99(1991).  
 RN [2]  
 RP REVISIONS.  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-96210632; PubMed-8633055;  
 RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 chromosome localization (9q24).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.  
 RC TISSUE-Glial tumor;  
 RX MEDLINE-97321050; PubMed-9177781;  
 RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,  
 RA Wiche G.;  
 RT "Plectin transcript diversity: identification and tissue distribution  
 of variants with distinct first coding exons and rodless isoforms.";  
 RL Genomics 42:115-125(1997).  
 CC -!- FUNCTION: PLECTIN IS PROPOSED TO PLAY A ROLE IN CROSS-LINKING  
 INTERMEDIATE FILAMENTS, TO INTER-LINK INTERMEDIATE FILAMENTS  
 WITH MICROTUBULES AND MICROFILAMENTS, AND TO ANCHOR INTERMEDIATE  
 FILAMENTS TO MEMBRANES, TO THE PLASMA MEMBRANE AS WELL AS TO THE  
 NUCLEAR MEMBRANE. PLECTINS HAVE THE ABILITY TO SELF-ASSOCIATE AND  
 FORM NETWORKS THAT STABILIZE THE CYTOPLASM.  
 CC -!- SUBUNIT: A TETRAMERIC STRUCTURE IS PROPOSED WHERE THE TWO CHAINS  
 IN ONE MOLECULE ARE PARALLEL TO ONE ANOTHER AND THE TWO MOLECULES  
 ARE ARRANGED ANTIPARALLEL AND OVERLAP TO A GREAT EXTENT.  
 CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE  
 PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN  
 SKELETAL MUSCLE AND LOWEST IN THYMUS.  
 CC -!- SIMILARITY: TO DESMOPLAKIN AND TO BULLOUS PEMPHIGOID ANTIGEN.  
 CC -!- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59601; CAA42169.1; -;  
 DR EMBL; U96274; AAC53209.1; -;  
 DR EMBL; U96275; AAC53210.1; -;  
 DR EMBL; U96276; AAC53211.1; -;  
 DR PIR; A39638; A39638.  
 DR PIR; S21876; S21876.  
 DR HSSP; Q01082; 1AA2.  
 DR InterPro; IPR001101; -;  
 DR InterPro; IPR001715; -;  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectin\_repeat; 21.  
 DR PROSITE; PS00019; ACTININ.1; FALSE\_NEG.  
 DR PROSITE; PS00020; ACTININ.2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton;  
 KW Actin-binding; Alternative splicing.  
 FT DOMAIN 1 400 ACTIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 1 71246 GLOBULAR.  
 FT DOMAIN 71247 2648 CENTRAL FIBROUS ROD DOMAIN R.  
 FT DOMAIN 71247 1474 R1 COILED COIL.  
 FT DOMAIN 1475 2648 R2 COILED COIL.

FT	DOMAIN	1475	1686	R2-A SUBDOMAIN (POTENTIAL).
FT	DOMAIN	1727	1927	R2-B SUBDOMAIN (POTENTIAL).
FT	DOMAIN	1967	2124	R2-C SUBDOMAIN (POTENTIAL).
FT	DOMAIN	2298	2489	R2-D SUBDOMAIN (POTENTIAL).
FT	DOMAIN	2490	2648	R2-E SUBDOMAIN (POTENTIAL).
FT	DOMAIN	2786	4620	GLOBULAR.
FT	DOMAIN	787	930	COILED COIL (POTENTIAL).
FT	DOMAIN	1105	1140	COILED COIL (POTENTIAL).
FT	DOMAIN	1241	1307	COILED COIL (POTENTIAL).
FT	DOMAIN	1378	1434	COILED COIL (POTENTIAL).
FT	DOMAIN	1472	2760	COILED COIL (POTENTIAL).
FT	DOMAIN	1727	1897	INTERACTION WITH VIMENTIN AND LAMIN B.
FT	DOMAIN	1727	1897	6 X TANDEM REPEATS OF MOTIF A.
FT	REPEAT	2786	4620	MOTIF A (APPROXIMATE).
FT	REPEAT	3114	3444	MOTIF A (APPROXIMATE).
FT	REPEAT	3445	3779	MOTIF A (APPROXIMATE).
FT	REPEAT	3780	4024	MOTIF A (APPROXIMATE).
FT	REPEAT	4025	4367	MOTIF A (APPROXIMATE).
FT	REPEAT	4368	4620	MOTIF A (APPROXIMATE).
FT	DOMAIN	2845	4597	6 X 19 AA REPEATS.
FT	REPEAT	2845	3015	1.
FT	REPEAT	3173	3343	2.
FT	REPEAT	3504	3674	3.
FT	REPEAT	3839	4008	4.
FT	REPEAT	4082	4252	5.
FT	REPEAT	4427	4597	6.
FT	DOMAIN	4628	4643	4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT	DOMAIN	4645	4687	SER-RICH.
FT	VARSPLIC	1	37	MSQRLRPVPEGLSGKRTSSDNLVLAIVRASEGKK (IN
FT				ISOFORM 2).
FT	VARSPLIC	1	37	MSQRLRPVPEGLSGKRTSSDNLVLAIVRASEGKK (IN
FT				ISOFORM 3).
FT	VARSPLIC	1	51	MSQRLRPVPEGLSGKRTSSDNLVLAIVRASEGKK (IN
FT				ISOFORM 4).
FT	VARSPLIC	38	180	MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	VARSPLIC	52	180	MISSING (IN ISOFORM 4).
FT	SEQUENCE	4687 AA;	533527 MW;	9966CA71B929751 CRC64;
SO	SEQUENCE	4687 AA;	533527 MW;	9966CA71B929751 CRC64;

Query Match 2.5%; Score 317; DB 1; Length 4687;  
 Best Local Similarity 18.3%; Pred. No. 0.0002;  
 Matches 545; Conservative 398; Mismatches 1096; Indels 932; Gaps 129;  
 Qy 10 QGAFSTEGSRYPHSHVQVTFNTRHQEF-----AVPDYRSSHLEVSQASQLLQOQQQ 62  
 Db 1546 EAAVDAQOQK---RSIQEELQHLRQSSEAEIQAKQAQVEAAERSMRIEEIRVVRQLQE 1602  
 Qy 63 QQLRRPSLLSEFHGSDRPOE---RRTSYEPFHPGPSVDHDSLSKRRPRLEQVSDSH 118  
 Db 1603 TTERQGGAEDELQALRAAEAEAKKQAOE-----EAERLRQVQDESQ 1648  
 Qy 119 FQVSAAYPLVPLVHPLPEGLRASADAKDPATGKGKEAPSSPTISGQPCGDQNASPSKLSK 178  
 Db 1649 RKRQAEAEAL-----RVKAEAEARE-----KQAL-----QALDELKQA 1685  
 Qy 179 BELTQSDMRVDREIAKVEQIILKKKKQOOLEEAEAKPEPEKPVSPPPVEQKHRSIVQ 237  
 Db 1686 EEAERWLQCAEAEARARQVQVALETAQRAEAEVQLQSKRPSFAEKTAQLERTLQEEHVITQ 1745  
 Qy 238 IYDENRKKAEAEHAKIFEGLPKVELPLYNPSDFKVVHENIKTNQVNRKKLILFFKRN 297  
 Db 1746 LREAERRAQQA-----EAERAEAEAELEWRQLKANEALRLR----- 1786  
 Qy 298 HARKOREQKICQVQDQLMEAEWEKKVDRIENPNRRKAKES-----KTREYVEKOFPEIRKQ 353  
 Db 1787 QAEEVAQOKSLAQADA-----EKQKEAEAREARRRGKAEQAVRORELAEQ---ELEKQ 1838  
 Qy 354 E-----QOERFQVQRGAG-----LSATITARSEHS----- 381  
 Db 1839 QLTEGTAQORLAEEQELIRLRAETTEQGEHQRLLEELARLQHEATAATQKROELEAELA 1898











FT CARBOHYD 951 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1679 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2054 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2244 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2362 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3030 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 349 MISSING (IN ISOFORM V1).  
FT VARSPLIC 1309 MISSING (IN ISOFORM V2).  
FT VARSPLIC 349 MISSING (IN ISOFORM V3).  
FT CONFLICT 348 P -> R (IN REF. 2).  
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 2.5%; Score 314.5; DB 1; Length 3358;  
Best Local Similarity 19.6%; Pred. No. 0.00017;  
Matches 457; Conservative 282; Mismatches 803; Indels 795; Gaps 119;

Qy 582 KGRITRSTNFA-AAASAAAAAATEEP--PPPLPPPPPEPISTE-PVETSRW-----T 629  
Db 349 KONISEATTIEMNLAETSSPLSKEPHMVPDRATVPVPLATELPIFTTHPPAGNIYNS 408  
Qy 630 BEEMEVAKKGLVEGRNMAALAKMVGYSKSEAQCKNFYKRRHNLNLLQOHKQKTSRK 689  
Db 409 EOKSVVYSQ-----ALTGRLAYESPITTRNTINSWDLNDSL-----ASGSG 449  
Qy 690 PREERDVQC---RSVASTVSAQDEDEDIEASNEENPDSSEVAVKPSEDS----- 738  
Db 450 PLGMPDISEIKEELRSTVTSQHATGSOAVITDQTHSVSQIEQIEVGPLVTSMEIT 509  
Qy 739 -----ENATSRGTEPAVELEPTE--TAPSTSPLAVPS----- 771  
Db 510 NHISLKEPKNKTPYESTE--VILEHTEMPTVSASPELATTSYGTFLREDDREDRTL 567  
Qy 772 -----TKPAEDESVEQVND--SISAETAEQMD-----VDQOEHS 804  
Db 568 TVRSQDSTRVFSQIPVITVSKTSED--TYSQGLDLESISTITMLGTDRLSDLDKEP 626  
Qy 805 ABEQSVCD-----PPATKADS---VDVEVRVPENHASKVGDNT-----KERDLORASE 851  
Db 627 KTNKGVTEDEFGSQPTTTPPSQHLTEVELLPYSQDTSVREGISTVIYPSLQTDVTOGRE 686  
Qy 852 KVE-PRDE---DLVVAQOINARPE-----PQSONDSSATCSADEV 889  
Db 687 RTETPRPELKDPYTVDEI-----PEKVTKDPFIKTEVFSGMPLSTSSSESVERTESV 742  
Qy 890 D-----CEPERQR-----MFPMDSKPSLLNPTGSL----- 915  
Db 743 SPALTIELTKPTEARDVEEMTTLTLETDVTKSDKDVTRVHLTHSTLVNVEVTVSKWP 802  
Qy 916 ---VSSPLKPNPLDPOLOHRA-AVIPP-----VSCTPCNIPGIPVSG-YALYQRHI 964  
Db 803 GDEDNSTSKP---LPSTEHAFTKLPVPLSTINGCKDKKEIPSFDTGGGEYTLFPDGT 858  
Qy 965 KAMESALLEQRORQOQIDLECRSSTSPCQTSKSPRE-----WEVLQAPAPQLITNL 1018  
Db 859 KPLEK---VSEEDIASGELATVTFHTSTSIGSAEKSASGEPTTGRFLPTTSTEDOVINAT 916  
Qy 1019 PEGVRL-----PTTRTPPPPLIPSKTTVASEKPSFTMGSSISQGTGTVL-TSHNOAS 1073  
Db 917 AEGSALGEDTASKPLTGPFFVHTSDV-----EELAFVNYSSTOE--PTIYVDISH----- 966  
Qy 1074 YTOETPKPSVGSISGLPROQESAKSATLPYIKOEFSPRSQSOPEGLVRAHQEGVVR 1133  
Db 967 -----TSPLSI-IPKTEWSVLETSPV-LEDEILLGKSDQDILEQTHLEATMSPGALR 1015  
Qy 1134 GTA---GAIQEGSITRGTPTTSKISVESI-----PSLRGSITOGT---P 1170  
Db 1016 TTGVSQGETQEEPTGSPPTFTSTAVMAKETTAFAEEGEGSTYTPS--EGLMTGSRVP 1074

Qy 1171 ALPOTGIPTALVKGSISRMPIE-DS-----SPEKGREEAASKGHVIEYEGKSG----- 1217  
Db 1075 GLETFPVT-SYPPGALTQOEVEDMTVWTLNSTRTPTVVSSTSEVIEYAEAGSSPTFAS 1133  
Qy 1218 -----HI-----LSYDNI-----KNAREGTRSPRTAHEISL----- 1243  
Db 1134 TLRPFQTHVQOLMEETEEGKKASLDVTLGSLGFEPERATELPRKFPPTSDISVFTAIDS 1193  
Qy 1244 -----KRSYESVEGNIKOGMSRESPSVAPLEGLICRALPRGSP 1282  
Db 1194 LHRTPPLSPSSFTTEQORVFEESSEKTDGILPGESVQHPVTVTLIDIVAMKT----- 1247  
Qy 1283 HSDLKERT-----VLSGSIQMGTPRATTEFEDGLKYPKQIK-----RESPP 1324  
Db 1248 ESDIDHMTSRPPVTPQTPRPSVVERKTTSTQELSTSPAACTKFHPDINVIIEVRENKT 1307  
Qy 1325 IRAFEGAITKGKPYDGIITTIKEMGRSIIHEIPRODILTOESRKTPEVVOSTRPI--IEGSI 1382  
Db 1308 GRRLSDMIVSGHPID-----SEKKEEPCSEETDPLHDLFAEI 1345  
Qy 1383 SQGTPIKF-----DNNSGQSAIKHNKSLITGPS-KLSRCMPPLLEIYENIKVVE- 1431  
Db 1346 LPELPDSFEIDIYHSEDEDEGDCV--NATDVTTPTSVOYINGKQLVTVPKDPEAAEA 1403  
Qy 1432 -RGKYEDVKAGETVRSRHTSVVSSGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSQYN 1490  
Db 1404 RRGQYESV-----APSQN-----FPDSSATDI---HQF 1428  
Qy 1491 TMSGSPMMNKTSDVTTPPNKSTNHERKSTLT--PTQRESIP--AKSPVGPVDVPVWSHP 1546  
Db 1429 ILA-----ETESSTMQFKKKEGTELETTWKRPETYPETPDHVSGEPDVFVFTLS--- 1479  
Qy 1547 FDPHHRGTADEVVWSHLPTQLDPAM--PEHRALDPAAAYLFOROLSPTPGYSQYO-L 1603  
Db 1480 ---SHDKT---TRWSESITESSNLENPVH-----KQKPVLPFFEESSGE 1520  
Qy 1604 YAMENTRQ-----TILNDYITSQOMQVNL---RPDVARGLSREQ 1640  
Db 1521 GAIEQASOETILSRATEVALGKETDQSTLSTSSILSSSVSVNLEERPLTLTGISQDTE 1580  
Qy 1641 PL-----GLPYATRGID-----LTNNMPTTILVPHPGGTSTP- 1673  
Db 1581 SMSTIESWEITPSTQTKFSESSAPIEGSGEYENKKNIFNMVTDLPQRPDPTDLSPL 1640  
Qy 1674 PMDRI-----TYIPGTQITFPPRPYNSASMP-GHPTHLAAAAS-----AERERERE 1721  
Db 1641 DMSKIMITNHHIYIPAT-----IAPLSKLPSPARPTTVWNSNSTSEWSDSKSFEGRKK 1695  
Qy 1722 KERERE-----RIAAASSD-LYLRPGSEQPCPGS-----HGVRSPSPSVR 1762  
Db 1696 KENEDEGAVNAHQGEVRAATERSDHLTLTPELESSNVDASSDLATWEGFILETPTT-E 1754  
Qy 1763 TOETMLOORPSVFGTNGTSTVI--TPLD-----PTAQLRIMPLPAGGPSISQGLPASRY 1814  
Db 1755 SEKEMANSTP--VFRETIGVANVEAQPFHESSSSHPRVOEELTTLGNNPPLSTDLGSDA 1813  
Qy 1815 NTAADALAALVDA-----AASAPQMDVSKTESKHEAARLEENLRSRAV 1860  
Db 1814 STGMELITASLFTLDLSETKVKKELPSTFSPSVEIS-----SSFEPTGLTPTSTVLDIEIA 1869  
Qy 1861 SEQOQLKQKTELEV-----KRSVQCLYTSAPPKQPQPHSSVWYSEAGKDKGPPPKS 1913  
Db 1870 GVMSQTSQKTLISISGKPTSQSGVRDLYT--GPPMGE---DFSGDSEYPTVSYPTMK- 1923  
Qy 1914 RYEEELTRGKTTITAAANFIDVITRQIASDKDARENGSSQSSSSSSSHRYETPSDAI 1973  
Db 1924 --EFTVMGG-----SD-DEVRVDTQTSSIPTTSDNIYVPVPSKG 1961  
Qy 1974 EVISPASSAPPQKQKLO-----TYOP-----EV 1996  
Db 1962 PDSTVASTAPPWEVEMSSAEGGSQLASVRSSVGPVPLAVDIFSCTESPYFDEEFEEV 2021



QY 1997 VKANAENDPT-----ROYEGLHRRFQESPPSQOQLPP-SSQAE 2037  
Db 2022 AAVTEANERPTVLTAAAGNTVDLTENGYIEVNTSMSLDFPOTMEPSKLWSKPEVNLKQ 2081  
QY 2038 GMQVPRTHRLITLADHICIIITODFARNOV-----SSQTPQPPPTST-----FONS 2084  
Db 2082 EIGR-----ETVTEKAQGQKTFESLHSSFAPEQITLQSLIETEFQTS 2126  
QY 2085 PSALVSTPVRTKTSNRYSPESQAQSVHRRPGSRVSPENLVKSRGSRPGKSPERSHVS 2144  
Db 2127 DYSMLTT-LKYIYTNKEVEEGMSIAHSTPG-----PGKIDLESYTH 2169  
QY 2145 EYEPISPQVPPVHVEKODSLLLLSQGAEPAEORNDARSFGSISYLPFFTKLENTSPM 2204  
Db 2170 -----PEAPGRSHFSATALTESGA-----AR-----SVLMSDSSTQEEESIKL 2208  
QY 2205 ----VKSKKQEIFRKLNSGSGSDMAAQPCTEIFNPAVTISGVSRSRSHGSPADPASN 2260  
Db 2209 FQKGVKLTNKSNDLSFGSLGSGGALPLPTTSV-NL-----TDMKQIITSLYAEFTSHMES 2264  
QY 2261 LGLEDIIRKALMGFDDKVEDH-----GVVMSQPMGVVPGTANTSVVTSGETRR 2309  
Db 2265 LG-----TSILG-----DKMEDHERMEDVSSNEVRMLISKIGISQSDSTEALDTLSHTGT 2315  
QY 2310'EEG-----DPSPH-----SGGVCKPK-----LISKNSRKSK-----SP 2338  
Db 2316 EPTTSTLFFVKMLDLERSPKQDPGSGGRKPKTHRPQTMGSLINSSEASAEAGATSP 2375  
QY 2339 ---IPGQYL-----GTERSSVSSVHS-EGDYHROTPGNAWEDRPSGTSTQ 2382  
Db 2376 TAFUP-QTYSVEMTKHFAPESSQPSDFNVNSGEGSEVD-----LDLVYTSQTQ 2426

## RESULT 29

SMFL\_HUMAN  
ID SWFL\_HUMAN STANDARD; PRT; 1902 AA.  
AC O14497; G9UP21;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotated update)  
DE SWI/SNF-RELATED, MATRIX-ASSOCIATED, ACTIN-DEPENDENT REGULATOR OF  
DE CHROMATIN SUBFAMILY F MEMBER 1 (SWI-SNF COMPLEX PROTEIN P270) (B120).  
GN SMARCF1 OR CLORF4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RX MEDLINE=20221560; PubMed=10757798;  
RA Dallas P.B., Pacchione S., Wilsker D., Bowrin V., Kobayashi R.,  
RA Moran E.;  
RA "The human SWI-SNF complex protein p270 is an ARID family member with  
RT non-sequence-specific DNA binding activity.";  
RL Mol. Cell. Biol. 20:3137-3146(2000).  
RN [2]  
RP SEQUENCE OF 1-1175 FROM N.A.  
RX MEDLINE=98094256; PubMed=9434167;  
RA Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;  
RT "Molecular cloning and expression of a novel human cDNA containing CAG  
RT repeats.";  
RL Gene 204:71-77(1997).  
RN [3]  
RP SEQUENCE OF 1-1132 FROM N.A.  
RA Takeuchi T., Misaki A.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS DNA NON-SPECIFICALLY.  
CC -1- SUBUNIT: PART OF THE SWI-SNF COMPLEX.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, THYMUS, PROSTATE,  
CC TESTIS, OVARY, SMALL INTESTINE, COLON, AND PBL, AND AT A MUCH  
CC LOWER LEVEL IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL  
CC MUSCLE, KIDNEY, AND PANCREAS.

CC -1- SIMILARITY: CONTAINS 1 ARID DOMAIN.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1132  
CC ONWARD DUE TO A FRAMESHIFT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; AF265208; AAF75765.1; ALT\_INIT.  
CC EMBL; AB001895; BAA23269.1; ALT\_FRAME.  
CC EMBL; AB024075; BAA83073.1; ALT\_SEQ.  
CC EMBL; AB024059; BAA83073.1; JOINED.  
CC EMBL; AB024060; BAA83073.1; JOINED.  
CC EMBL; AB024061; BAA83073.1; JOINED.  
CC EMBL; AB024062; BAA83073.1; JOINED.  
CC EMBL; AB024063; BAA83073.1; JOINED.  
CC EMBL; AB024064; BAA83073.1; JOINED.  
CC EMBL; AB024065; BAA83073.1; JOINED.  
CC EMBL; AB024066; BAA83073.1; JOINED.  
CC EMBL; AB024067; BAA83073.1; JOINED.  
CC EMBL; AB024069; BAA83073.1; JOINED.  
CC EMBL; AB024068; BAA83073.1; JOINED.  
CC EMBL; AB024070; BAA83073.1; JOINED.  
CC EMBL; AB024071; BAA83073.1; JOINED.  
CC EMBL; AB024072; BAA83073.1; JOINED.  
CC EMBL; AB024073; BAA83073.1; JOINED.  
CC EMBL; AB024074; BAA83073.1; JOINED.  
CC MIN; 603024; -.

CC InterPro: IPR001606; -.  
CC Pfam: PF01388; ARID; 1.  
CC DNA-binding; Nuclear protein.  
CC DOMAIN 96 99  
CC POLY-GLN.  
CC POLY-GLN.  
CC POLY-SER.  
CC ARID.  
CC GLN-RICH.  
CC W->A: PARTIAL LOSS OF DNA-BINDING  
CC ACTIVITY. WHEN A.A. 713 ALSO MUTATED,  
CC COMPLETE LOSS OF ACTIVITY.  
CC Y->A: PARTIAL LOSS OF DNA-BINDING  
CC ACTIVITY. WHEN A.A. 690 ALSO MUTATED,  
CC COMPLETE LOSS OF ACTIVITY.  
CC D->G (IN REF. 1).  
CC V->M (IN REF. 1).  
CC Q->S (IN REF. 2).  
CC G->GG (IN REF. 3).  
CC P->S (IN REF. 2).  
CC P->L (IN REF. 2).  
CC MANPPQVGSQMCP -> NGQYATSGWVRDVS (IN REF.  
CC 2).

CC MUTAGEN 713 713  
CC CONFLICT 27 27  
CC CONFLICT 51 51  
CC CONFLICT 349 349  
CC CONFLICT 368 368  
CC CONFLICT 374 374  
CC CONFLICT 393 393  
CC CONFLICT 489 502  
CC CONFLICT 634 634  
CC CONFLICT 924 924  
CC CONFLICT 1016 1016  
CC CONFLICT 1033 1033  
CC SEQUENCE 1902 AA; 205946 MW; 9B70A46F81062EAA CRC64;  
Query Match 2.5%; Score 312; DB 1; Length 1902;  
Best Local Similarity 17.4%; Pred. No. 0.00011;  
Matches 340; Conservative 260; Mismatches 678; Indels 672; Gaps 85;

QY 604 TEEPPPPPLPP-----PPEISTEPVETSRWTEEMEVAKKGLVHGHRNWAIAKWGTK 657  
Db 19 SQQQGPPSDPQQGHGYPGQPGYQTPQRYPMYVQGRASAMGGLSYTQQIIPPYQGGPSG 78  
QY 658 SEACKNFYFNKRRHNLNLLQHKQKTSRKRPRERDVSCVESVASTVAQEDIEDIAS 717  
Db 79 YGQCGQTPTYNQSQSPHP-----QQQQPPYSQPPSPQTPHAQ-----PSYQQQPPSQPPQ 127



QY 718 NEEENPESEVEAVKPSDESDENATSRGNTPEVALEPTTETAPSTSPSLAVPSTK-PAE 776  
Db 128 LQSSQPPYSQPPSPHQQSP-----APYPSQSTTQOHQSPQPPYSPQQAQSPYQ 178  
QY 777 DESVETQVNDISLAEAT-EQMDVDQOEHSAEBSVCDPPPPATKADSDVDVVRVVENHASK 835  
Db 179 QOQPOOPAPSTLSQAAAYPQOSQOSQOTAYSQORFPFPQELSDQSFQSOASSAPMTSS 238  
QY 836 VEGDNTERDLRASEKVEPRDELDVAAQINAOPEPEQSDNDSSATCSADEDDVDEPER 895  
Db 239 KGG-----QEDMNLQ-----SRPSSLPLSGS-----IDD----- 265  
QY 896 QRMFPMDKPSLLNP-----TGSILVSSPLKPNPLDLPQLOHRAAVIPPMVSTPCNIPGT 952  
Db 266 ---LPMGTE-GALSPGVSTGSSSQGQSNPAQSPFSPHTSPHLPIRGSPS--PVGS 319  
QY 953 PVSQYALVQRIKAMHESALLEBOQROEQIDLECRSTSCPGTSGSKSNRREWEVLQAPH 1012  
Db 320 PAS-----VAQSRSGPLSPAAPVPGN----- 340  
QY 1013 QLITNLPEGVRLPTRPRPPPLIPSPKTTVASEKPSFIMGSGISQGT-----PCTYLTS 1068  
Db 341 -----MPPRPP-----SGQSDIMHPSMNQSSIAQDRGYMORN 373  
QY 1069 HNOASYTOETP-----KPSVGSISLGLPRQESAKSATLPIYIKQEEFSPRSONSQPEG 1121  
Db 374 PQMPQYSSPQPGSALSPPQSGGQIHTGMVSYQQNSMGS-----YGPQGGQYGPQ 424  
QY 1122 LLVRAQHEGV-----RGTAGIQEISITRGPTTSKISVESPSLRGSIQTGTPALPQ 1174  
Db 425 GYRQPNYNALPNANYPAGMAGGIN-----PMGAGGQMHGQPGIPP 466  
QY 1175 TG-IPTALVKGISIRMP-----IEDSSPEKG-----REEAASKGHVIEGKS 1216  
Db 467 YGTLPGRMSHASMGRNYPGNMAMPQVSGCMCPPPGGNRKTQETAVAMHV----- 520  
QY 1217 GHILSYDNKNAREGTSRPTAHEISLRSYVESVGNIKQMSMRSPVSA-----PLRGL 1272  
Db 521 ----AANSIQNRPG-----YPNNGGMMGTGPPYGGINSMAGM 557  
QY 1273 ICRALPGSPSHDLKERTVLSGSI-----MGG-----TPRATTESFEDGLKY 1314  
Db 558 I---NPOGPPYS-----MGGTMANNAGMAASPENMGLGDVLTLPATKMNKADGTPK 607  
QY 1315 PKQIKRSPPIRAFEAGATK-----GKPYD-----GTTIKEMGRSHE 1353  
Db 608 TESKSKSSSTTTNEKITLYLGGEPRKMWDRYLAFTEEKAMGNTLNPVARGPLD 667  
QY 1354 IPQDILTQESRKTPVYVQSTR-PIEGSISQGTPIKFDNNSGOSAIKHVYKSLITGPK 1412  
Db 668 LYRLVSVKEIGGLTVNKNKKRELATNLNVT-----SSSAASLLKQYIQCLYAFCK 723  
QY 1413 LSRGMPLEIVPENIKVVERKYEYEDVKAQEVVRSRHSYVSSGFSVLRLSTLHEAPKAQ-L 1471  
Db 724 IERGEDP---PPDI-----FRAADSKSQ-----PKIQPP 750  
QY 1472 SPGIYDDTSARRTVSYONTMSRSPMNRTSDVTIPNPKSTINHERKSTLTPQRESTPA 1531  
Db 751 SPAGSGSQGGQTQTQSTSSNAEG-----GDLKPTTAPSTHSQ-----IP- 791  
QY 1532 KSPVPGVDPVVYSHSPFDPHHRGSTAGEYVWSHLPTQLDPAMPFFHRALDPAAAYLFRQ- 1590  
Db 792 --PLPGMSRSVGIQDAFNDGSDS-----TFQKRN 820  
QY 1591 -LSTPGY-PSQYQLYAMENTROTILNDYITSQQMQLNRPDPVARGLSPREQP----- 1641  
Db 821 SMTNPGYQPSM-----NT-----SDMGRMSYEPNKPDPYGSWRKAPGSDPFMSSGQ 868  
QY 1642 ---LGLPYPATRGI-IDLTNMPPTILVPHPGTSTPMDRITVIP-----GTQITPPR 1691  
Db 869 PNGMGDPYSRAAGPGLGNVAMGRQHPYCG-----PYDRVTEPGIGEGNMSTGAPQ 923  
QY 1692 PYSNASMPGHPHTHAAAAAER 1751

Db 924 P-----NLMPSNP-----SGMYSPSRYPPOOQOQQRH-----DSYGNQFSTQ-GTPSG- 968  
QY 1752 GYVRSPPSVTQETMLQOORPSVFGTNGTSTVITPLDPT-----AQLRIMPLPAG- 1801  
Db 969 ----SPFSSQGT---TMYQOQOONYK-----RPMQGTGPPAKRHEGEMYSVPYSTGQ 1014  
QY 1802 GPSISQGLPASRYNTAADALAAALVDAASAPQMDVSKTKESKHEAARLEENLRSRAAVS 1861  
Db 1015 GQOQOQQLPPAQPASQO-----QAQPSQOQDYNYOYGNAYPAT----- 1055  
QY 1862 EQOLEQKLTLEVERKRSVOCLYTSAFPGKPOPHSSVYVYSEAGDK-GPPPKSYEEELR 1920  
Db 1056 -----ATAATER-----PAGGQONQFPFQF---GRDRVSAPPCTNAQ--- 1091  
QY 1921 TRGKTTTAAFNIDVITITRQIATSKDKDARERSQ---SSDSSSLSSHRYETPSDAIEVIS 1977  
Db 1092 -----NMPQMGPGIQAESAQVQAQGTMMQGRNDMTYNYANRQ-----S 1130  
QY 1978 PASSPAPPOEKLQTYQPEVVKANQANDPTQYEG--PLHHRPQOESPS---PQOQLP 2031  
Db 1131 TGSAPQGPAYHGVNRTDEMLHTDQAN-----HEGSWPSHGTRQPPYGPSPAPVPMTRPP 1185  
QY 2032 PSSQAEGMGQVPRTHRLITLADHICQIITQDFARNQVSSQTPQOQPTSTFQNSPVSALYST 2091  
Db 1186 PSN-----YQPPPSMQNHIPQVSSPA 1206  
QY 2092 PVRTKTSNRYSPESQAOVHHQRPGSRVSPENLVDKSRGSRPGKSPERSHVSSPYEPIS 2151  
Db 1207 PLPRPMENRTSP-SKSPFLH---SGMKM-----QKAGPPVPASHIAPAPVQD--- 1249  
QY 2152 PPQVPVVHEKQDLSLLLSORCAEPAEORNDARSQSYLSYLPSTFKLENTSPMKVSKKOE 2211  
Db 1250 -----PMIRDTTFPGSV-----EATQPVLKQRRRL 1276  
QY 2212 IFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSSRSHGFADPASNLGLDIIKAL 2271  
Db 1277 TMKDGT-----PEAWRVMMSLKSLGAE-----STWALDTI----- 1308  
QY 2272 MGSFDDKVEDHGVVMSQPMGVVPGTANTSV-----VTSGTTRREGDPSPHSGVC 2322  
Db 1309 ----NILLYDDNSITMNLSQLPGLLELLVEYFRCLIEIFGILKEYEYVDPGQRT----- 1360  
QY 2323 KPKLKSNRSKSPKIPGQYGLTER-----PSSVSVSHSEGDYHIQTPGAWEDRPSS 2377  
Db 1361 ---LLDGRSKVSSPAPMEG--GEEEEEELGPKLEEEVEEVEENDEEIAFSGDKRPAS 1415  
QY 2378 TGSTQ---FPYNPLTMRMLSTPTPTPIACA 2404  
Db 1416 ENSEKLSKFDKLPVKIVQKNDFVVDCS 1445  
RESULT 30  
MOZ\_HUMAN  
ID MOZ\_HUMAN STANDARD; PRT; 2004 AA.  
AC Q92794;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MONOCYTIC LEUKEMIA ZINC FINGER PROTEIN (ZINC FINGER PROTEIN 220).  
GN ZNF220 OR MOZ.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96376968; PubMed=8782817;  
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,  
Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,  
Horsman D., Mielman F., Volin S., Watmore A.E., Housman D.E.;  
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses  
a putative acetyltransferase to the CREB-binding protein.";



RL. Nat. Genet. 14:33-41(1996).  
CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL  
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE  
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.  
CC -!- SIMILARITY: CONTAINS 2 PHD ZINC-FINGER DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U47742; AAC50662.1; -.  
CC HSP: G06031; 4GBQ.  
CC MIN: 601408; -.  
CC InterPro: IPR001965; -.  
CC InterPro: IPR002717; -.  
CC Pfam: PF01853; MOZ\_SAS; 1.  
CC Pfam: PF00628; PHD; 2.  
CC  
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;  
KW Nuclear protein.  
FT ZN\_FING 206 256 PHD-TYPE 1.  
FT ZN\_FING 259 313 PHD-TYPE 2.  
FT DOMAIN 371 379 POLY-SER.  
FT ZN\_FING 538 560 C2HC-TYPE.  
FT DOMAIN 788 1478 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 788 798 POLY-GLU.  
FT DOMAIN 989 995 POLY-GLU.  
FT DOMAIN 1019 1026 POLY-ARG.  
FT DOMAIN 1069 1078 POLY-GLU.  
FT DOMAIN 1147 1150 POLY-LYS.  
FT DOMAIN 1225 1228 POLY-GLU.  
FT DOMAIN 1267 1275 POLY-GLU.  
FT DOMAIN 1294 1302 POLY-GLU.  
FT DOMAIN 1411 1414 POLY-GLU.  
FT DOMAIN 1593 1597 POLY-SER.  
FT DOMAIN 1647 1701 GLN/PRO-RICH.  
FT DOMAIN 1647 1650 POLY-GLN.  
FT DOMAIN 1651 1656 POLY-PRO.  
FT DOMAIN 1659 1668 POLY-PRO.  
FT DOMAIN 1668 1675 POLY-PRO.  
FT DOMAIN 1690 1696 POLY-PRO.  
FT DOMAIN 1707 2004 MET-RICH.  
FT SITE 1546 1547 BREAKPOINT FOR TRANSLLOCATION TO FORM  
FT MOZ-CBP.  
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;  
  
Query Match 2.58; Score 310.5; DB 1; Length 2004;  
Best Local Similarity 17.18; Pred. No. 0.00013;  
Matches 358; Conservative 276; Mismatches 640; Indels 823; Gaps 87;  
  
Qy 55 QLLQQQQQLRRPRLSEPHGSDRQERTSYEPF---HPGFS-----PVDHD 102  
Db 318 KLLQKAAQIKRRYTNIG--RPNRLKKQNTVSKGPFVKVTKGGRKKKIYLSQSA 375  
Qy 103 SLESKRLRLEQVSDSHFORVSAAYL-----PLVHPLPEGLRASADAKDPAFG----- 150  
Db 376 SSSSEGYLERIDGLDFCRDSNVSLRNFNKKTKGLDGLTKFTFPDGRK--ARGEVDY 433  
Qy 151 -----CKHEAPSPISGQPCGDDONASPKLSKEELI-----QSMRDVREIAK-VQOI 199  
Db 434 SEQYIRKRGNRKSSSDWPT-DNQDGDCKQENELFGSQEIMTEKDMELFDIQEA 492  
Qy 200 LKLLKKQQLLEAAKPEPEKPV-----SPPEVE----- 229  
Db 493 LQ-----KVGVTGPPDQVRCPSVIERGKYEIHTWYSSPYQEYSLRPLKYLCEFL 544

Qy 230 --QKHSIVQIYIDENRKKAEAHKIFGLGPKVPELPLYNQPSD--TKVYHENI----- 279  
Db 545 KYMSRTILO---QHMKKCGWFPPANEIYRKNINISVFVDGNVSTIYQCNLCILAKLF 600  
Qy 280 ---KT-----NOVMRKKLIILFFKRRNHARKOREQKICQYD---OLMEA 317  
Db 601 LDHKTLYYDVEPFLFYVLTQNDVAGCHLVGVFSKEKHQ---QKYNVSCIMILPQ 652  
Qy 318 WEKK---VDRIENPRRAKESKTREYVEKQFPETRKQREOQERFORVQORCA--- 367  
Db 653 YQKGYGRFLID-----FSYLLSKRE-----GQAGSPEPK 682  
Qy 368 ---GLSATIARSEHSEIIDGLSEQENNEKQRLQSVIPPMFMFADQRRVKFTNMNGL 423  
Db 683 LSDGLRLSYMAYWKSVILECLYHQNDKQISIKKLSKLTGICPDITSLHLHRLMD--- 738  
Qy 424 MEDPMKYKDRQPMNVWTDHEKEIFKQFIQHPKNFGLIASYLERKSV-PCVLYYYITK 482  
Db 739 -----FRSQDFV-----IIRREKLIQ--DHMAKLIQLNLRPVDVDPCLRWTPVIV 781  
Qy 483 KNEYKALVRRNYGKRGRNQOIARPSQEEKVEEEDKAEKTEKKKEEKDEE----- 536  
Db 782 SN-----SVSEEEEEEEEGEENEPEQCOERELEISVG 814  
Qy 537 -----EKDEKEDSKENTKEKDIDGTAEETEERQATPRGRKTANSQGRRKGRITR--- 587  
Db 815 KSVSHENKEQDSYSVESEKKEPVMAVSSTRLSKQVLPHDSLPAANSQPSRRGRGKRNK 874  
Qy 588 -----SMTNEAAAAASAAAAATEPPP-----PP 610  
Db 875 TQERFGDKSKLLEETSSAPQEQYEGEGEKSEATQEQYTESQQLVASEEQSDGKPD 934  
Qy 611 LP-----PPPEPISTEPETS---RWTEEMEVAKKGLVEHG----- 644  
Db 935 LPKRLSEGVEPWGQLKKSPALKCRLTGSERLPRYSEGDRAYL-RGSESESEEEE 993  
Qy 645 -RNWAAIAKVMGTKEAOCKNFYFNKRR-----HN-----LDNL----- 679  
Db 994 PESPRSSPPILTKPTLKRKKPFLHRRRRVRKRHHNSVYVTTTSETTEVLDFEDES 1053  
Qy 680 -----QOHKOKTSRPRE-----ERDVSOCESVASTVSAQEDDEA 716  
Db 1054 SERPMPRLPTFEDEEEEDENELPFREYRRLSSQDLRCQSSSKRKSDEEDES 1113  
Qy 717 SNEENPEDSEVEA-----VKPSDSDPENAT-----SRGTEPAV 751  
Db 1114 DDADDTPLKPVSLLRKRVKNSPLEPDTSTPLKKKGKPKSRKPIHWKKRPGKPGF 1173  
Qy 752 ---ELEPTTAPSTSPSLAVPT--KP---ADESVETQVNDNISAEATAEQMDVDQOE 802  
Db 1174 KLSREIMPVSTQACVIEPIVSIKAGRKPKIQESEETVEPKEDMPLPERKEEEMQAEA 1233  
Qy 803 HSAEGSVCD-----PPPATKADSVDFVVRVYPENIHASKVEGDNTKERDLDRASEKE 854  
Db 1234 EEAEGEEDAAASVEVPAASPADSNSPETETKEPE-----VE-----EEEEKRVSEQR 1284  
Qy 855 PRDEDLVVAQINARPEPQSDNDSSATCSADEVDGSEPERQRMFPMDSKPSLLNPTGSI 914  
Db 1285 QSEE-----EQOLEEPEPEEEDAAATAQNDHDDADEDDGHLESTKKE----- 1331  
Qy 915 LVSSPLKNPLDLPOLQHRAAVIPPVMVSCPTCPTIGTPVSGYALYQRIKAMHESALLE 974  
Db 1332 LEEQPTREDVKKEPGVQE-----SFLDANMQS-----REKIDKKEETELDS 1373  
Qy 975 EQRQREQIDLECRSSSTSPCGTSKSPNREWEVLQAPAHQLITNLPEGVRLPTTRTPPP 1034  
Db 1374 EEEQ----- 1377  
Qy 1035 PLIPSSKTTVASEKPSFTMGGSISQGTPTGTYLTSHNOASYTOETPKPSVGSISLGLPRQ 1094  
Db 1378 ---PSHTSVVSEQ---MAGED-----DHEEDSHYKE----- 1404  
Qy 1095 ESAKSATLPYIKQEFSPRSQNSQPEGLLVRAQHEGVVRGTAGATQEGSITRGTPTSKIS 1154







Db 468 SJSN---PNPIDPSSMORAYAAALGLPYMNQPTQLOPQVPGQQPAQPPAHO---QMRTLN 521  
Qy 422 GLMBDPMKV-----YKROFMNVWTDHEKEI---FKDKFIQHKNFGLIASYLERKSVPD 473  
Db 522 ALGNPNMSVPAGGITTOQPPNLSALPTSLSGATNPLMDGNSNGISL---STIP- 577  
Qy 474 CVLYYYLTKKNENYKALVRRNYGRRKRNQOJARPSQBEKVEEKEE-----DKAEKTEK 527  
Db 578 -----TAAPPSSGTVRKWHEHVTDLRSLHVKLVQAIFPTPPDPAALKDR 623  
Qy 528 KEEE-----KKDEEKEDEKEDSKENTKE--KDKIDGTAETEEREQATPRGRKTANSGR 580  
Db 624 RMENLVAAYKVEGDYTESANDEYHLLAEKLYKTOKELEER-----RTLHKQG- 676  
Qy 581 RKGRITRSMTEAAAAAATEEP-----PPLPPPPPEPISTEPVETSRWTEEMEVEV 635  
Db 677 -----ILGNQFALPASGAQPPVIPPQAOSVRPNGLPLPLPVNRQVSGQMNSFNMSL 728  
Qy 636 AKGLVHGRNMAIAKM---VGTKSEBAQCKNFYNYKRRHNLNLLQKHQK--TSRKPR 691  
Db 729 GNVQLPQAPMPRAASPMNHVSQMNMAVPGMAISPSRMPQPPNMGTHANNIMAQAPT 788  
Qy 692 EERDVSCQ-----SVASTVSAQEDIEDIASNEBE-----NPEDSEVAV 731  
Db 789 QNQFLPQNFSSGAMSVNSVGMGPAAQAGVSGQBPGEAALPNPLNMLAPQASOLPC- 847  
Qy 732 KPSEDSP-----ENATSRGNT--EPAVELEPTTETAPSTSPSLAVPST 772  
Db 848 PVTQSPHLPTPPASTAAGMPSLQHTAFGNTPPQAPATQPTVPSSGOTPT-PTPGS 906  
Qy 773 KPAEDESVEQVNDISIAETAQMDVDQOEHSAEESGVCDP-----PPAT-- 817  
Db 907 VPS---AAQTOSTPTVQAAQAQV--TPQPTPVQPPSVATPQSSQOQPTPVHTQPPGTP 962  
Qy 818 ---KADSDVDVVRVENASKVEGDNTERDLDRASEKVEPRDEDLVAQAQINAORPPEQS 875  
Db 963 SQAASIDNRPTSTVTSATSSQOPGPDVPMLEMTQVOTD-----AEPEPTE 1013  
Qy 876 DNDSSATCSADEVDGPERORMPMDSKPSLLNPTGSLVSSPLKNPLDLPOLQHRAA 935  
Db 1014 SKGEPRSEMDEEDLGSSQVKEETD-----TTEQKSEPMVEEKKPEVK 1057  
Qy 936 VIPPMVSCPTCNIPGTPVSGYALYQRIKAMHESALLEQROEQROQIDLECRS--STSPC 994  
Db 1058 V-----EAKEEENSNDTASQSTSPS 1079  
Qy 995 GTSK---SPNREWEVLOPAPHOLLITNLPQVRLPTTRTPRPPLIPSSKTTTVAERPSF 1051  
Db 1080 QPRKKIFKPEELQALMPTLEALYRQDPES--LPFRQPDVQLLGIPIDYDIVKNPMDLS 1137  
Qy 1052 IMGGSISQ---TPCTY-----LTSHQASVTOETPK-----PSVGS 1086  
Db 1138 TIKRKLDTGQYQEWYQYVDDVRLMNFNANLYNRKTSRVYKFCSLAEVFEQIDPVWQSL 1197  
Qy 1087 SLGLPROQESAKSATLPYIKOEFSRPSQNSQPEGLLVRAQHEGVRGTA--GAIQESIT 1145  
Db 1198 GYCCGRKYEFSPQLCCYKQLCTIPR-----DAAYVSYQNRVHFQCKFTEIQENVT 1251  
Qy 1146 RGTPTSISVESIPSLGSIQTGTPALPOTGIPTALVKGSIRMPDIETSSPEKGREAA 1205  
Db 1252 LG-----DDPSQPTTISKDQFEKKNDTL--DPEPFVDCKECG 1288  
Qy 1206 SKGHVIEGKSGHILSYDNI-----KNAREGTRSPRTAHEISIKRSYESVEGN----- 1253  
Db 1289 RKMHCIC-----VLHYDIIWPSGFVDCNLUKTKGRPRKENKFSAKRLQTLRLGNHLEDR 1342  
Qy 1254 IKQGMMSRESVPASPLEGLICRALPRGPHSD-----LKERTVLGSGSIMQGTTPRATT 1305  
Db 1343 VNKFLRQNHPEAG-----EVFVRVASSDKTVEVKPGMKSRFVDSGENSESFYRTK 1395  
Qy 1306 ESFDEGLKYPKQIAKRESPPIRAFGATKPKPYDGIITIKEMGRSINE-----IPRQDILT 1361  
Db 1396 ALF-----AFE-----EIDGV-DVCFPGMHVQDITALIAPHQ----- 1425

Qy 1362 QESRKTPEVVOSTRPIIEGS--ISQGTPIKFDN--NSGOSAIAKH-----VRSIL- 1407  
Db 1426 -----TQGVYISYLDISIHFRRPCLRTAYVHEILIGLYEVVKLVYVT 1469  
Qy 1408 -----TGPSKLSRGMPPLEIVPENIKVVERGK-----YEDV--KAGE 1442  
Db 1470 AHIWACPPSGDDYIFHCHPPDQIPKPKRLQEWYKKMLDKAFARIINDYKDFKQANE 1529  
Qy 1443 TVSRHSTSV-----SSGPSVLRSILHEAPKAQLSPGIVDDTSARETPVSYQ-----NT 1491  
Db 1530 ---DRLTSAKELPYEGDFWPNVLEESIKELEQEEERKKEETAASSETPEGSGDSKNA 1586  
Qy 1492 MSRGSPMMNRT--SDVT-----IPPNNKSTNHERKSTLTPQRES-----IPA 1531  
Db 1587 KKKNNKTKNKSSISIRANKKPSMPNVSNLSQKLYATMEKHKEVFFVIHLHAGPVIST 1646  
Qy 1532 KSPVPGVDVVVS-----HSPF-----DPHHRGSTAGEVVVS----- 1562  
Db 1647 QPPIVDPDPLLSCLMDGDRDAFLTARDKHWFSLSLRKSKWSTLCMLVELHTQGDREYV 1706  
Qy 1563 -----HLPTQ-----LDPAHPFHAL-----DPAAYL 1586  
Db 1707 TCNECKHHVETRWHTCTCEDYDLICINCYNTKSHTHKMVKWGLGLDDGSGSGEPOSKSPQ 1766  
Qy 1587 FQROLS-----PTPGYPSQYLYAM 1606  
Db 1767 ESRLSITORCISLVHACOCRNANCSLPSCOKMKRVYQHTKGCKRTNGGCPVKQLLAL 1826  
Qy 1607 -----ENTROTILNDYITISQMQ-----VNLRPDVARGL--SP 1637  
Db 1827 CCYHAKHCQENKCPVPFCLNFKHNVRQOQIQHCLQQAQLMRRMATMTRNVPOQSLSP 1886  
Qy 1638 REQPLGLPYPATRGIDLTNNPPIILVPHPGTSTP--PMDRI--TYIPGTQITFP- 1690  
Db 1887 TSAPPGTP-----TQOPSTPQTPPPAQPQPSVNNMSPAGFPNVAQTPTTIVISA 1936  
Qy 1691 -RPNYSASMSPGHPTHLAAASAEERERERERERERERERERERERERERERERERERER 1749  
Db 1937 GKPTNQVPAPPAPPPAPPAVEAARQIEREAQOQOHLR---ANINNGMPFGRDGMGTGP 1993  
Qy 1750 SHGVVSPSPSVRTQETMLQORPSVFGTNGTSTVITPLDPTAQLRIMPLPAGGSPISQGL 1809  
Db 1994 SQ-----MTPVGLNVRPNQVSGPVMSM-----PPQWQQAIPQOQP--MFGM 2036  
Qy 1810 PASRYNTAADALAAALYDAASAPQMDVSKTESKHEARLEENLRSAAYVSEQOOLQOK 1869  
Db 2037 P-----RPVMSQAAQAAVAGPRMNPVQPNRSISPSA-LQDLTLTKSPSPQOQ--QQ 2086  
Qy 1870 TLEVEKRSVQCLYTSSAF-----PSGKPOP--H-----SSV 1898  
Db 2087 VLNILKSNPQLM---AAFIKORTAKYVANQPMQPOGLOQSQPMQPOGHHQOQPSLQNL 2143  
Qy 1899 VYBAGKDK-GPPPKSRVYEBELTRGTITTAANFID-----VIITROI- 1941  
Db 2144 NAMQAGVPRPGVPPPPQPMAGGLNFGQO---ALNIMNPGHNPNTMNMNPQYREMYRROLL 2199  
Qy 1942 --ASDKDARERGSSOSSSSSSLS--HRYETPSDAIEVISPAS--SPAPQEKLOIYQ 1993  
Db 2200 QHOQO 2253  
Qy 1994 P-----EVVYKANO---AENDPTRO---YEGPLHHYRPOQESPSPOQOLPP 2032  
Db 2254 PIQSSMGQMAAPMQLGQMGQPGGLGADSTNIOQALQOQILQOQOQOQOQOQOQOQOQOQO 2313  
Qy 2033 SSOAEGMGQVPRHRLITLADHI--CQIITQDFARNQVSSQTPQOQPTSTFQNSPSALVS 2090  
Db 2314 SPQOQHMLSGOQO-----ASHLPQOQIATS--LSNQVRSAPVQSPRPQ-----S 2355  
Qy 2091 TPVRTKTSNRYSPESQAQSVHH-----QRPGRSVSPENLVKSRGSRPGKS----- 2136  
Db 2356 QPPHSSPSPIPOQ---PSPHHVSPQTGPHPLGLAVTWASSMDQGHGHNPEOSAMLQJLN 2412



Qy	2137	-PERSHVSSE	2145
		:	
Db	2413	TPNRSALSSE	2422

RESULT 32

ID	CBP_HUMAN	STANDARD;	PRT;	2442 AA.
AC	Q92793; Q16376; O00147;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	CREB-BINDING PROTEIN.			
GN	CREBBP OR CBP			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxID=9606;			
[1]	SEQUENCE FROM N.A.			
RP	MEDLINE=97321049; PubMed=9238046;			
RX	Sobulo O.M., Borrow J., Tomek R., Reshmi S., Harden A.,			
RA	Schlegelberger B., Housman D., Dodgett N.A., Rowley J.D.,			
RA	Zeleznik-Le N.J.;			
RT	"MLL is fused to CBP, a histone acetyltransferase, in therapy-related			
RT	acute myeloid leukemia with a t(11;16)(q23;pl3.3).";			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).			
[2]	SEQUENCE FROM N.A.			
RP	MEDLINE=97321049; PubMed=9177780;			
RX	Giles R.A., Petrij F., Dauwerse H.G., den Hollander A.I.,			
RA	Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,			
RA	Dodgett N.A., Peters D.J.M., Breuning M.H.;			
RT	"Construction of a 1.2-Mb contig surrounding, and molecular analysis			
RT	of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome			
RL	16p13.3.;"			
RL	Genomics 42:96-144(1997).			
[3]	SEQUENCE OF 1-405 FROM N.A.			
RP	MEDLINE=96376968; PubMed=8782817;			
RX	Borrow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,			
RA	Chaganti R.S.K., Clivin C.I., Distche C., Dube I., Frischauf A.M.,			
RA	Horsman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;			
RT	"The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses			
RT	a putative acetyltransferase to the CREB-binding protein.;"			
RL	Nat. Genet. 14:33-41(1996).			
-1-	FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO			
CC	PHOSPHORYLATED CREB PROTEIN, ACTING AS A COACTIVATOR. CBP AUGMENTS			
CC	THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF			
CC	CAMP-RESPONSIVE GENES.			
-1-	SUBCELLULAR LOCATION: NUCLEAR.			
CC	DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLLOCATION			
CC	T(8;16)(P11;P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)			
CC	INVOLVING CBP AND MLL.			
-1-	DISEASE: DEFECTS IN CBP ARE THE CAUSE OF RUBINSTEIN-TAYBI			
CC	SYNDROME (RTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL			
CC	ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION			
CC	AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.			
-1-	SIMILARITY: CONTAINS 1 BROMODOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collabor			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstat			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for comm			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annou">http://www.isb-sib.ch/annou</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
EMBL	U47741; AAC51770.1; -			
DR	EMBL; U85962; AAC51331.1; -			
DR	EMBL; U89354; AAC51339.1; -			
DR	EMBL; U89355; AAC51340.1; -			
DR	MM; 600140; -			



















CC . CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS  
CC AT A DEFINED SITE, PCE/THE, WITHIN THE HCF REPEAT (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D45419; BAA08258.1; -  
DR HSSP; P02751; 1FNA.  
DR InterPro; IPR001777; -  
DR Pfam; PF00041; fn3; 2.  
KW Nuclear protein; Repeat.  
FT REPEAT 44 89  
FT REPEAT 93 140  
FT REPEAT 148 194  
FT REPEAT 217 265  
FT REPEAT 266 313  
FT REPEAT 313 366  
FT DOMAIN 1010 1448  
FT REPEAT 1010 1035  
FT REPEAT 1072 1097  
FT REPEAT 1101 1126  
FT REPEAT 1157 1182  
FT REPEAT 1295 1320  
FT REPEAT 1323 1348  
FT REPEAT 1358 1383  
FT REPEAT 1423 1448  
FT SEQUENCE 2090 AA; 214942 MW; E495EBB1F2385E17 CRC64;

Query Match 2.4%; Score 300.5; DB 1; Length 2090;  
Best Local Similarity 19.5%; Pred. No. 0.00033;  
Matches 338; Conservative 193; Mismatches 641; Indels 565; Gaps 69;  
QY 966 AMHESALLEQRORQEOIDLECRSTSPCGT-----SKSPNEWEVLQAPAPQL 1014  
DB 388 ATADSYLLQKY---DIPATATATPTNPVSPANPPKSPAPAAAAPAPVQLPQVG 444  
QY 1015 ITNLEPGVRLPTRPPPLPSSKTV-----ASEKPSF--IMGGSISQGTPTLT 1067  
DB 445 ITLVQAAAAPPSTTTIQLVTPVGSISVPTAARAGVPAVLKVTGPQTGTPLVTR 504  
QY 1068 SHNQASTQETPKPSVSGISLGLPRQESAKSATLPYIKOEESPRQNSQPEGLLVRAQ 1127  
DB 505 PAGQAGKAPVTVTSPLASVRMVVPTQ--SAQGTVI-----GSPNQMSGMAALA- 550  
QY 1128 HGVVRCGTAGAIQSGITRGTPTSKISVESIPS-----LRGSIQTGPALPOTGIPTAL 1182  
DB 551 -----AAAAATQ-----KIPPSAPVLSVPAGTTIVKVVPTGTTLPAT----- 592  
QY 1183 VKGISRMPEDSSPEKREAAKSHVIEGKSHILSYDNIKNAREGTRSPRTAHE-- 1240  
DB 593 VKVASSPVVSNPATMLKTAQAQVGVSVSSA-----ANTSTRPIITVHKSG 639  
QY 1241 -ISLKRYSY---SVEGNKMGMSRESVPSAPLEGLICRALPRGSPHSDUKERTVLSGSI 1296  
DB 640 TTVTAQAQVVTTVVGGVTKTITLVKSPISVPGGSALISNL--GKVMVSVQTKPVQTSV 697  
QY 1297 MQGTPRATTSFEDGLKYPKQIKRESPIPAFEGAIK-----GKPYDGIITIKGRS 1350  
DB 698 ---TQASTGPV-----TQIQTQKPLPA--GTILKLVTSADGKPTTIITTTQASGAG 745  
QY 1351 -----IHEIPRODILTQESR-----KTPVQVOSTPII 1378  
DB 746 SKPILGISSVSPSTTKPGTTTIKTIPMSAIIHQAGATGVTSTPGIKSPITTIITKMT 805  
QY 1379 EGSISQGTPIKFDNNSSQSAIKHNKVSILITGPKSLRGMP--LEIIVP-ENIKVVERGKY 1435  
DB 1379 APADSLNDPSTESCLN-----ELASAVPST-VALLPSTAT 1807

DB 806 SGT---CAPAKIITAVPKIATGHGQGVTVQVLKAGPQPGAILRTPVMSGVRLVTPVT 862  
QY 1436 EDKAGETVRSRHTSVV--SSGPSVLSTLHEAPKAQLSPGIDYDTSARRTPVSVQNTMS 1493  
DB 863 SAVKPAVT-----TLVKGTTGVTLTGTVGTSTSLAGAHSTSLATPITLTGTTIA 917  
QY 1494 RGSMMNRTSDVTIPIPPNKSTNHERKSTLT-----PTQRESIPAKSPV- 1535  
DB 918 -----TLSSQVINPATAVSAOATLTAAAGGLTTPITMQPVSQPTQVTLITAPSGVE 970  
QY 1536 -----PG-VDPVSVSHSPDPHRRSTAGEV 1559  
DB 971 AQPVHDLPLVILASPTTEQPTATVTIADSGQGVQVTVTLVCSNPPCETHTGTT----- 1026  
QY 1560 YWHLPTQLDPAMPFHRALDPAAAAAYLFQRLSPTPGVPQYQLYAMENRTQTLINDYIT 1619  
DB 1027 -----NTATTTVANLGGHP-OPTQVQFVC---DRQEAASLVT 1061  
QY 1620 SQMQVNLDPDVARGLS--PRE-OPLGLPYPATRGIIDLTMN-----PPTILVPHPGGT- 1670  
DB 1062 SAVGQON--GNVVRVCSNPPCETHTGTTTATATATSNMAGOHGCSNPP--CETHETGTT 1117  
QY 1671 -----STPPMDRITYIPG---TQITPPRPYNASMSPGHPT 1704  
DB 1118 STATTAMSMGTGOORDTRHTSSNPTVVRIIVAPGALERTQGTVPKQ---COTQOANMT 1173  
QY 1705 HLAASAAER 1764  
DB 1174 NTMTVQATR-----SPCAGPILLRPSVALEAGNHSFAFVQLALPSVRV- 1217  
QY 1765 ETMLQRPVSFQGTNGTS-----VITPLDPTAQLRIMPLPAGGPSISQGLP 1810  
DB 1218 -----GLSGPSNKDMPTGHOLETVHTVTTTPTTALSIM--GAGELGTARLIP 1263  
QY 1811 ASRYNT-----AADALAAALVDAASAPQMDVSKTKESKHEAARLEENLSRASA 1860  
DB 1264 TSTYESLOASPSSTMTMTALEALLCPSATVTVQVCSNPPCET-HETG--TTNTATTSNAG 1320  
QY 1861 SEQOOLEKTLVEKRSVQCLYTSSAPPSGPKPQPHSSVVYSEAGK----- 1906  
DB 1321 SAQVCSNPPCETHTGTTTATATATATATATATATATATATATATATATATATATAT 1373  
QY 1907 -----KGPPPKSRYEEELTRGKTITTA---ANFIDVITRIADSK---DA 1947  
DB 1374 TGTMTSVSVGALLDPATPSHCTLESGLVAVSVTVSOAGATILLASFPQVCSNPPCET 1433  
QY 1948 RERGS---QSSDSSSLSHRYETP--SDAIEVIS----- 1977  
DB 1434 HETGTTTATTTVTSNMSSNODPPPAASDQGEVSVTSSGDSANITSSSGITTTVSTLPRAV 1493  
QY 1978 -----PASSPAPPQEKLOTYQP-----EVVKANQANDPTROYEGPLHY 2017  
DB 1494 TTVTQSTPVPGPSVNPISLSTETFCALTSEVPPIPATIVTANTTETSDMPFSAVDILQPP 1553  
QY 2018 RPOQESPSPOOLPP-----SSQAGMGQVPRTHRLITLADHICQIITODF 2063  
DB 1554 EELQVSPGPRQOLPRLQLQASSTPLMGESSEVLASQTPELQAAVDL----- 1601  
QY 2064 ARNOVSQTPQOQPTSTFQNSPSALVT-----PVRTKTSNRYSP--SQAQSVHHQ 2113  
DB 1602 --SSTGDFSSGQETS-----SAVATVVVQPPPTQSEVDQLSLPOELMAEAQAGTTT 1653  
QY 2114 RPSGRVSPENL-----VDKSRGS 2132  
DB 1654 LMVTLPEELAVTAAEAQAATAEAQALATQAVLQAAQAVMAGTGPMDTSEAAA 1713  
QY 2133 PGKSPERSHVSSEYEPISPPQVFWHEKQDLSLLLSQR-----GAPBQRN---DAR 2183  
DB 1714 AVTQAEGLHLSAEGEG-QATTIPVLVQAEALVQVQVQVQVQVQVQVQVQVQVQVQV 1772  
QY 2184 SPGISYLPSTFTKLENTSPVWKKBQIFRKLNSGGSDMSDAAAQPGTIFNLPVTT 2243  
DB 1773 APADSLNDPSTESCLN-----ELASAVPST-VALLPSTAT 1807















```
DR EMBL; D17635; BAA04538.1; JOINED.
DR EMBL; D17636; BAA04538.1; JOINED.
DR EMBL; D17637; BAA04538.1; JOINED.
DR EMBL; D17639; BAA04538.1; JOINED.
DR EMBL; D17640; BAA04538.1; JOINED.
DR EMBL; D17641; BAA04538.1; JOINED.
DR EMBL; D17642; BAA04538.1; JOINED.
DR EMBL; D17643; BAA04538.1; JOINED.
DR EMBL; D17644; BAA04538.1; JOINED.
DR EMBL; D17645; BAA04538.1; JOINED.
DR EMBL; D17646; BAA04538.1; JOINED.
DR EMBL; D17647; BAA04538.1; JOINED.
DR EMBL; D17648; BAA04540.1; JOINED.
DR EMBL; D17634; BAA04540.1; JOINED.
DR EMBL; D17636; BAA04540.1; JOINED.
DR EMBL; D17637; BAA04540.1; JOINED.
DR EMBL; D17639; BAA04540.1; JOINED.
DR EMBL; D17640; BAA04540.1; JOINED.
DR EMBL; D17641; BAA04540.1; JOINED.
DR EMBL; D17642; BAA04540.1; JOINED.
DR EMBL; D17643; BAA04540.1; JOINED.
DR EMBL; D17644; BAA04540.1; JOINED.
DR EMBL; D17645; BAA04540.1; JOINED.
DR EMBL; D17646; BAA04540.1; JOINED.
DR EMBL; D17647; BAA04540.1; JOINED.
DR EMBL; D17648; BAA04540.1; JOINED.
DR EMBL; D17552; BAA04490.1; JOINED.
DR EMBL; M26684; BAA04811.1; JOINED.
DR PIR; A32642; A32642.
DR HSP; P05412; IFOS.
DR InterPro; IPR000075; -.
DR Pfam; PF02029; CALDESIGN.
DR PRINTS; PR01076; CALDESIGN.
KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW Repeat; Alternative splicing.
FT DOMAIN 26 199 MYOSIN AND CALMODULIN-BINDING.
FT REPEAT 251 390 10 X 13 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 251 265 1.
FT REPEAT 266 278 2.
FT REPEAT 279 291 3.
FT REPEAT 294 306 4.
FT REPEAT 309 321 5.
FT REPEAT 324 336 6.
FT REPEAT 337 349 7.
FT REPEAT 350 362 8.
FT REPEAT 363 375 9.
FT REPEAT 378 390 10.
FT DOMAIN 523 580 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 622 632 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 612 644 STRONG ACTIN-BINDING.
FT DOMAIN 674 680 CALMODULIN-BINDING.
FT DOMAIN 726 752 WEAK ACTIN-BINDING.
FT DOMAIN 39 46 POLY-ARG.
FT DOMAIN 539 542 POLY-GLU.
FT DOMAIN 556 559 POLY-GLU.
FT MOD_RES 485 485 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 597 597 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 682 682 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 688 688 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 711 711 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 717 717 PHOSPHORYLATION (BY CDC2).
FT VARSPIC 1 24 MDDFERRELRQRKEERLEAER -> MISRSYCRQNL5

Query Match 2.4%; Score 298.5; DB 1; Length 771;
Best Local Similarity 20.9%; Pred. No. 0.00012;
Matches 182; Conservative 124; Mismatches 334; Indels 231; Gaps 34;

QY 293 FKRNRHAKRQEQKICQRYDQLMEAWKVKVDRIE---NNPRKAKESKTR-----E 340
DB 4 FERRELRQRKEERLEAERL--SYQRNDDDEEAARERRRRRRAQRQERLQKEEGDVSGE 61
QY 341 YFEKQFPEIRKQEQEQRFQVQGRGAGLSATIRSEHSEIIDGLSEQENNEKQMRQL 400
RT
```

```
DB 62 VTEKSEVNAQNSVAEEETKRSTDDEAALLERLARREERQRKLQELAEKQKEDPTITDG 121
QY 401 SVIPPMFDAEQRRVKFINMGL-----MEDPMKYVKDQFMNWTDEH- 444
DB 122 SLSP-----SRRENNVEENITGKEEKVETROGCEIEETETVTKSQRNWRDGE 176
QY 445 --KEIFKDKFIOHPKNFGLIASYLERKSVDCVLYYYLTKKKNYKALVRRNYGKRGRN 502
DB 177 EGKKEEKDEEEKPEVP-----TEENQDVAV-----EKSTDKSEVVETKTLAVNA 225
QY 503 -----QIARPSOEKVE-EKEEDKAEKTEK-----EKEKDEEKEDEKESKENT 548
DB 226 DTNMLEGEQISITDAADKEEAEKEREKLEAEKAEKAEKAEKAEKAEKAAE 285
QY 549 KKKIDGTAE--TEERQA--TPGRKKTANSOGRKGRITRMTNEAAAAAASAAAAATE 605
DB 286 EERERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAE 345
QY 606 EPPPLPPPPPEPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 665
DB 346 R-----AKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAE 380
QY 666 YFNKRRHNLNLLQHKOKTSRKPREDVSOCEVASTVSAQEDIEDA---SNEEN 722
DB 381 ----KRAAEKARLEAKLKKKKMEKK--AQEEKAQANLLRKQEDKAEKAEKESL 434
QY 723 PE----DSEVAVKPSEDS-----PENATSRGNTPEPAVELEPTETA 760
DB 435 PEKLOPSTKKQVKONKOKEKAPKEEMKSVDRKRGVPEQKANGERE----- 482
QY 761 PSTSPSLAVPSTKPAEDSVETQVNDISAEQMDVQDQHSABEGSVCDPPATKAD 820
DB 483 -----LITPKLSTENAFGRSNLKGAAAGSEKLEKQEAVE----- 523
QY 821 SYDVEVRPENASKVEGNTKRDLDRASEKVEPRDEDLVAAQINAORPEQSDNDS 880
DB 524 -LD-ELKRRERRRRIIEEQKKQEEAEKIREBEKKRMKEETERRAE-----A 574
QY 881 ATCSADEVDGEPERMPMDKPSLLNPTGSL-----VSSPLKP----- 922
DB 575 AEKROKVPEDGVSEKKPF-----KCFSPKGSSLUKIEERAEFLNKSOKGKPAHTTA 628
QY 923 --NPLDLPOLHRAAVIPPMVS--CTPCNIPITGTPVSGYALYQRIKAMHESALLEQRQ 978
DB 629 VVSKIDSLRLEQYTSVAVGNKAAPAKPAASDLVPVAGV---RNIKSMWEG----- 677
QY 979 ROEQIDLECRSTSPCGTSKSPNREWEVLQAP-----HOLIINLPGVRLPTRPT--RP 1032
DB 678 -----NVFSPGGTG--TPNKETAGLVGVSSRINEWLTKTPGKNKSPAPKPSDLRP 727
QY 1033 -----PPPLIPSSKTTVASEK 1048
DB 728 GDVSGKRNLEWQSVKPKPAASSKVTATGKK 758

RESULT 37
SC16_YEAST
ID SC16_YEAST STANDARD; PRT: 2194 AA.
AC P48415;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MULTIDOMAIN VESICLE COAT PROTEIN.
CN SEC16 OR YPL085W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96017704; PubMed-7593161;
RA Espenshade P., Glimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
RT "Yeast SEC16 gene encodes a multidomain vesicle coat protein that
```



interacts with Sec23p.";  
J. Cell Biol. 131:311-324(1995).  
- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE  
ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND  
WITH THE CYTOSOLIC DOMAIN OF SCD4. COULD THEREFORE BE A  
CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION  
CAUSES A LETHAL SECRETION DEFECT.  
- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES  
WHICH BUD FROM IT.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U23819; AAC49088.1; .  
DR SGD; S0006006; SEC16.  
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.  
FT MUTAGEN 1058 L->S: IN SEC16-4; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1083 L->P: IN SEC16-3; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1088 L->P: IN SEC16-2; TS ACCUMULATION OF ER  
MEMBRANES.  
FT MUTAGEN 1230 W->R: IN SEC16-1; TS ACCUMULATION OF ER  
MEMBRANES.  
SQ SEQUENCE 2194 AA; 241613 MW; BB1E02D2AD4683E3 CRC64;  
-----  
Query Match 2.4%; Score 298; DB 1; Length 2194;  
Best Local Similarity 18.7%; Pred. No. 0.00043;  
Matches 474; Conservative 324; Mismatches 897; Indels 842; Gaps 121;  
QY 14 STEQRYPHSVQYTFPTTRHOEFVAVDYRSSHLEVSOASOL-LQQOQQOQLRRRPSLL 72  
DB 39 STINSFNDDSVNRTESDIASKSD--VPPV-SSSTNISPANETQLEIPDTQELHHK--LL 93  
QY 73 SEFHGSDRPQRRYSYEPFHGPGSPVDHDS-----LESKRPRLEQ 113  
DB 94 -----NDSQDHITADSNLDPSNIVEHDSVITOTKPMASQEYEETAHLSSRNPSLDV 147  
QY 114 V-----SDSHFQRYSAVPLVHLPLPEGLRASADAKDPAP---GGKHEA----- 155  
DB 148 VAGELHNNHETQKTAVS-----AVEEDSFNEEGEENHDSIIISLND 190  
QY 156 -----PS-----SPISGQPCG-----DDQ-----NASP 173  
DB 191 ATPSQYNHFLPSDGNLSPSELSSGDTPTHNVPLGTDKNEINDDEYCNDEKISLNANNVLP 250  
QY 174 SKLSKEELIQSDMRVDREIAKVEQOILKKKQOOLEEAAKPPPEKVPSPVPEQKHR 233  
DB 251 DELSREE-----DERLKEITHVSTEERKQDIADQETA---ENLFTSSTEPSSENKIR 298  
QY 234 SIQVLIIDENKKAEEAHKIEPGLGPKVELPLYNQPSDTKV-YHENIKTNQVMRKLLIF 292  
DB 299 N-----SGDDTSMLFQ-----DDESOKVPWEEDVKD----- 326  
QY 293 FKRRNHARKQEQKTCORYDQLMFAWEKKVDK---IENPRRKAKESKT---REYKEQ 345  
DB 327 FHNETNTNQ-----ESAPNTDQDRKGYEGNEALKKSECTAERSYSSET 373  
QY 346 PPEIKRQREQ-----ERFORVQOR--GAGLSATIAARSEHISEIIDGLSEQENN 393  
DB 374 SEDIFPHGDKQVQEQNDFTGKNINIENESQKLMGEGNHKLPLLSAEADIIE--PGKDIOQA 431  
QY 394 EKQMRQLS-----VTPPMFQ-----AEORRVKFINMGLMED-PMKYTKDRQFMNV 439  
DB 432 EDLFTQSSGDGLGEVLPWSTQKNADVTSKSQKHEDLFAASQNDKLPWEV-SQGEVSSG 490  
QY 440 WTDHEKEIFDKFIQHPKNF-----GLIASYLERKSVDPCLVLYIT-----K 482

DB 491 KTENSMQSTEKIAQKPSFLENDDDLLDDDDFLASSEEDTVPNTDNTNLTSPVEK 550  
QY 483 KNEYKALVRRNYKRRGR-----GRKKGRITRSMTEAAAAA-----NQIARPSOEKVEE 516  
DB 551 KASRYKPIFEEDAGMRQEQVHTNTTGTIVTPOQFHLTKTGLGTPNQOVSVNIVSPKPP 610  
QY 517 KEEDKAEKTEKEEKKBDE-----EKDEKEDSKENTKEKDKIDGTAEETEEQQA 567  
DB 611 VKDNRSNFKINEEKKKSDAYDFPLEIISSESKKHAKPVAVPTQRF-GSGNSFSLLDKP 669  
QY 568 TPRGRKTANSQ-----GRKKGRITRSMTEAAAAA-----ATEEP 607  
DB 670 IPQSRKGSNNRPPVPLGTQEPSSR--TNSAISQSPVYAFNPKYKIQLOQAQIS 727  
QY 608 PPPLP-----PPP-----EPISTEPVETSRWTEEMEVAKGLVHGRNMAAIAKVMGTGS 658  
DB 728 GMPLENTNIPPPALKVETTVSAPPTR-----ARGVSNAS 761  
QY 659 EAQCKNFYNTKRRHLDN---LLOQHKOKTSRKPREEDYVSOCEVASTVSAQEDDIE 715  
DB 762 VGSSASFGARATQGLNNGVPPVSPYQATINLP---TANKYAPVSPTV----- 808  
QY 716 ASNEEENPEDSEVAVKPSEDSPEN--ATSRGNETPAVELEP-----TTETAPSTPSLIA 768  
DB 809 ---QOKQYPSVVQNLGASAVNTPNFKVTHRGHTSSISYTPNQNEHASRYAPNTQQSQY 864  
QY 769 VPSTKP-----AESVETQVNDISISAEABQMDVDQEHSAEBSGVCDDP-----PA 816  
DB 865 VPYTSQVGPVAGNSYSQSTRSSYAVPMMPQA---QTSASIQPHANIQPTGILPLAPL 921  
QY 817 TKADSVDEVVRVPEHNSKVEGDNTERDLDRASEKVEPRDEDLVVAQOINAOPEPQSD 876  
DB 922 RPLDPLQATNL-QPRASNTAANSLPLANUPLAENILP---EIITHRATSSVAPROEN 977  
QY 877 NDSSATCSADEVDGEPERQRMFP-----MDSKPSLLNPTGSLIVSS---PLKPN 923  
DB 978 NPIK-----IDNEALLRQFPFIHWSAANKVYVAPPIDQSOYMISSIVQEIKT 1029  
QY 924 PLDLPLQHLRAAVIPP--WVSCPTCNIPIGTPVSGYALYQRIHAKMHESAL--LEEQROR 979  
DB 1030 PID-----QIIKPNMLKSPF-----GPLGSAKLKKDLTKWMTETIKSISENESS 1075  
QY 980 QEQ-----IDLECRSSTSPCGTSKSPNREWEVL---OPAPH-OLITNL----- 1018  
DB 1076 TDMTWIQLLEKMLNDKVNWKNSKLLYNSDELLMYLSQFPFNGDMIPNAYRDLINCMQVR 1135  
QY 1019 -----PEGVRLPTTRTPRPPPLIPSSKTTTVASEKPSFIMGSGISQ--TPG----- 1063  
DB 1136 LAFLOTGNHDEALRLAKSKROYAIALVGS---LMGKDRWSEVIQKLYEGTAGPNQDK 1192  
QY 1084 -----TYLTSNQASYTOETPKPSVSGSISLGLPROQESAKSAT 1101  
DB 1193 ELAHLFLLIFQVFGNSKMAIKSFYTNNETSQWASENKSIVAVALINIPENNED--PLL 1250  
QY 1102 LPYIKQE---EFSPRSQNSQPEGLLVRAHQHEGVVGTAGAQEGSITRGTPTSKTSVESI 1158  
DB 1251 IPPVVVLEFLIEF-----GIFL-----TKKGLTAAASTLFILGNVPLS----- 1287  
QY 1159 PSLRGSIQTGPALPOTGIPTAEALVKGISRMPIEDSDSPEKGRGEEAASKGHVYIEGKSGH 1218  
DB 1288 -----NEPVMADSDVIFE---SIGNMNTES-----ILWDEIYEV 1319  
QY 1219 ILSYDNINKNAREGTRSPRTAHEISLKRYSVEGNIKQGM-----SMRESVPV 1265  
DB 1320 IFSYDP-KFKGFSSTLPQKIYHASLLQ-----EQGLNSLGTKYTDYLSSSVVKLPK 1369  
QY 1266 SAPLEGLICRALPGSPHSDLKERTVLSGSIQCTPRTATESFEDGLKYPKQIKRESPP 1325  
DB 1370 KDILITINLTREL-----SEVASRLESNTGWLAKPKLSSVWQDLKSFNKTIGDD--I 1421  
QY 1326 RAFEGAITKGPYDITTIKEMGRSIEHPRQDILTOESRKTPEVVQSTRPIEGSISQ 1385



Db 1422 DALNKKNDKKVDFGFTP-----GSSANS-----STVDLTQTFPTF-----QAQV 1461  
Qy 1386 TPIKFDNNSGOSALKHNKVS---LITGSKLSRGMPLLEIVPENIKVYKRGKEDVRAGE 1442  
Db 1462 TSQSYVDFTALLHNAHVPSVLSHSPNSVKG-----LVEANLPYTHR--IGDSLOGS 1514  
Qy 1443 TVSRHTSVSSGSPV-----LRSTLHEAPKAQISPGIYDTSARRTPVSYONTMSRGP 1497  
Db 1515 PORIHTQFAAAEPOMASLRVRVDQHTNEKALKSQOILEKKSTAYTPQFGON---HSPV 1571  
Qy 1498 MNRTSDV-----TIPPN--KSTNHERKSTLPTTORESIPAKSPVPGVDVP 1541  
Db 1572 MEKSNVPSLFAFPAPPKLGTVPNSVSPDLVRRESIISTGSEFLP--PPKIGV--- 1626  
Qy 1542 VHSPPFDPHRRGSTAGEYVWS---HLPTOLDPAMP-----F 1574  
Db 1627 -----PTKANSOGSLMSPSVEALP--IDPVVPQVHETGYNDFGNKHSQKSWPEDES 1677  
Qy 1575 HRAIDPAAAYLFORQLSPT-----PGYPYOYQLYAME-----NTRQTI 1613  
Db 1678 HTSHDNSNAD---QNTLKDSADVTDETMDIEGPGFNDVKLLPMEPNHQPSTVNPQT 1734  
Qy 1614 LNDYITSOQMOVNLN-----PDVARGLSPREOPLGL-----PYPATRGIDLT 1656  
Db 1735 SDDIQLITNVVEGTDASKHNSLPSIENRSESEOPENISKSASAYLPSTGGLSLE 1794  
Qy 1657 NMPTILVPHPGGTSTPMDRITVPGTQITFPFRPYNSASMSGHPHTHLAAASAEER 1716  
Db 1795 NRPLT-----QDENSISSETVQSTVLPAGSISMEAKPISQVQDVPRVNN--KASKLVEOHM 1848  
Qy 1717 EREKERERERI-----AAASDLYLRPGSEOPGRPGSHGYVRKSPSV--RTQE 1765  
Db 1849 APPKPKSTDAKMNYSVPVPOSTAASAD-----GDESTILKTSPIAYARTHQ 1895  
Qy 1766 TML-----QORPSVFCQNGISVITPLDPTAQLRIMPLPAGGPSISQGLPASRYNTAADAL 1821  
Db 1896 AHASNPQYPLVNO--ANETASFELSESTSOAQ-----SNGNVAENRFSPIKK 1943  
Qy 1822 AALVDAASAPQMDVSTKESKHAAR--LEENLRSAVAASEQOQ-----LEQKTLVE 1874  
Db 1944 AEVVEKDTFQP--TIRKASTNQYRAFPLESDADKYNDVIEDESDDDNMSTDEAKNRKE 2001  
Qy 1875 KRSVOCLYTSAPFSGKP-QPHSSVSYSEAGKDGKPPKPSYEEELRTRGKTTITANFI 1933  
Db 2002 KKNVN--MKKETPSNKDIDDKSNGWGLKKDTG--DKVKYKAKL--GHKNTL---YY 2051  
Qy 1934 DVITITROIASDKDARERSQSSDSSLSHRYETPDSAEIVSPASSPAPQEKLOTYQ 1993  
Db 2052 DEKLKRW--NKDATEEKQ-----KITESSAPPPPIVIRKDDGG 2089  
Qy 1994 PEVYKANQAENDPTROYEGPLHRYPOQESPSQOQLPPSPSQAGCMGVPRTHRLITAD 2053  
Db 2090 PK-----TKPRSGPINN-----SLPPV----- 2106  
Qy 2054 HICQIITQDFARNVSSQTPOOPPTS---TFQNSPSPALVSTPVTKTSNRYSPSQASV 2110  
Db 2107 HATSVI-----PNNITGEPLPIKTSPS-----PTGPNPNNSPSS----- 2143  
Qy 2111 HHQRPGRKVPENLVDK 2127  
Db 2144 ----PISRISGVNLTSK 2156

RESULT 38  
MAP2\_MOUSE  
ID MAP2\_MOUSE STANDARD; PRT; 1828 AA.  
AC P20357;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE MICROTUBULE-ASSOCIATED PROTEIN 2.  
GN MAP2 OR MTAP2.  
OS MUS musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89083571; PubMed=3205744;  
RA Wang D., Lewis S.A., Cowan N.J.;  
RT "Complete sequence of a cDNA encoding mouse MAP2.";  
RL Nucleic Acids Res. 16:11369-11370(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89043973; PubMed=3142041;  
RA Lewis S.A., Wang D., Cowan N.J.;  
RT "Microtubule-associated protein MAP2 shares a microtubule binding  
RT motif with tau protein.";  
RL Science 242:936-939(1988).  
CC -1- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY  
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO  
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.  
CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; M21041; AAA39490.1;  
DR PIR; S06467; S06467.  
DR PIR; A40115; A40115.  
DR MGD; MGI:97175; Mtap2.  
DR InterPro; IPR001084;  
DR Pfam; PF00418; tubulin-binding; 3.  
DR PROSITE; PS00229; TAU\_MAP\_1; 2.  
DR Microtubules; Repeat; Calmodulin-binding (POTENTIAL).  
FT DOMAIN 1452 1472 CALMODULIN-BINDING (POTENTIAL).  
FT REPEAT 1662 1692 TAU/MAP MOTIF.  
FT REPEAT 1693 1723 TAU/MAP MOTIF.  
FT REPEAT 1724 1755 TAU/MAP MOTIF.  
SQ SEQUENCE 1828 AA; 198980 MW; 200BC59E360538CA CRC64;  
  
Query Match 2.3%; Score 296.5; DB 1; Length 1828;  
Best Local Similarity 19.2%; Pred. No. 0.00039;  
Matches 371; Conservative 226; Mismatches 669; Indels 667; Gaps 91;  
  
Qy 832 HASKVEGDNTKERDLDRASEKVEPRDELVVAQOINARPEPQSDNDSSATCSADEVDG 891  
Db 61 HRSQGYSDTKENGING-----ELTSADRETAEEVSARIVQ-----VVTAEAVAVLKG 108  
Qy 892 EPERQRMFPMDSKPSLLNPTGSLVSPKLPNPLDLQLOHRAAVIPWVSCPTCNPIG 951  
Db 109 EQEKEAQY-----KDPAALPLAAEETANLPSPPPSP----- 141  
Qy 952 TPVSGYALYORHIKAMHESALLEEQORQOEIDLECRSSTSPCGTSKSPNREWEVLPAP 1011  
Db 142 -----ASEQTATVEE-----DLTLAS-----KMEFFPEQEKFP 168  
Qy 1012 HQLITNLPVGLRPTTRP-----PPPLIPSSKTTVASSEKPSFIMGSGTSQGTG----- 1062  
Db 169 SSFAEPLDKG-EMEFKMPKSGEDFEHAALVPDTSKTPQDKKDLQGMGEGLPPVPFAQT 227  
Qy 1063 -GTYLTSNQASYTQETPKPSVGSISLGLPQOESAKSATLPYIKQEEFSPRSNQSEQ 1121  
Db 228 FGTNLEDRKQST-----EPSIVMPSTIGLSAEPPAPKPKDWFIEMPTESKDE----- 275  
Qy 1122 LLVRAHQGVVVRGTAGATQEGSITRGTPTSKISVESIPSLRGS-----ITQGTAL 1172  
Db 276 -----WGLAAPISPCPLT--PMREKDVLEDIPRWEGKQFDSMPSPFFGSGFTL 322  
Qy 1173 PQTGIPTEALYKGSISRMPIEDSSPEK-----GREAAKSHVYIEGKSGHI--LS 1221



Db 323 PLDTMKNKRVSEGPFPAPVFQSDDKVSLQDPSALATSKSSKDEEPLKADKADVDS 382  
Qy 1222 YDNKNAREGTRSPRTAEISLKRSEYVEGNKOGMSRSPSAPLEGICRALPRGS 1281  
Db 393 ISEVTTILGNVHSPVVEGVG-----ENISGEVTTTQOEKRETSAP--SVQEPFLTETE 435  
Qy 1282 FHSCLKERTVLS-----GSMQGTTPRATTESFEDGLKYP-KQIKRE 1321  
Db 436 PQTCLDEKSTVSEIEBAVEESLKLDRDKTGVQTSFQSFREDKQGOEHTIDELAQD 495  
Qy 1322 SPPTRAFGATKGPYDGITTIKEMGRSIIHEIPRODILLQESRKTPEVVOSTRPIEGS 1381  
Db 496 SFPI-SLEQAVT-----DAAMTSKTLGR-----VTSE-----PEAV-SERREIOGL 534  
Qy 1382 ISOGTPIKFDNNSGOSALKHNKVSLLITGPSKLSRGMPLEIVPENIKVVERGKYEDYKAG 1441  
Db 535 FEETADKNKLEGGASA-----TIAEVEMPFYD----- 563  
Qy 1442 ETVRSRHTSVVSSGSPVLSTLHEAPKAQLSPGIYDDTSARRTPVS-----YQNTM 1492  
Db 564 -----KSGMSK-----YFETSALKEDMTRSTELGSDYIELSD 595  
Qy 1493 SRGSPMMNRTSDVITPPNKNSTNHERKSTLTPTQRESIPAKSPVPGVDVYV-SHSPFDPHH 1551  
Db 596 SRGS-AQESLDTISPKNQ--HDEKELQAKASQSPPAQE--AGYSTLAQSYTPGHP-- 646  
Qy 1552 RGSTAGEVWVSHLPOLDPAMPFHR--ALDPAARAYLQROLSPTPGYPQSYQVLYAMENT 1609  
Db 647 -----SELPE--BFSPQOERMFTDPKV-----YGERKDLHSKNKD 680  
Qy 1610 RQTILNDY-----ITSQOMQVNL-----RPDVARG-----LSPREQPLGLPYPATR 1650  
Db 681 DITLSRSLGLGRSAEQRSKNSINLPMSCLSIALGFNFGCHDLSPLASDI---LTWTS 737  
Qy 1651 GIIDLTN--MPTTILVHPHGTSPTPPNDRIYIP-----G 1683  
Db 738 GSMDEGDDYLPPT-----TPAVEKMPCFPTESKEEDKAEQAKVTGGQTIQVETS 787  
Qy 1684 TQITFPPRPY--NSASMSGPHGTHLAAA-----ASAEREREREKERERERTAAASD 1735  
Db 788 SESPPFAKEYKNGTVMAPDPEMLDLATGRSLASVADAEVARRKSVSPSEAMLAES 847  
Qy 1736 LYLRPGSQP--GRPGSH-----GY-----VRSPSPSVRTQETMLQORPQSVFQGTNGTS 1782  
Db 848 SLPPVADESPTVVRPDSQLEDWGYCFVKNYTVPLSP--VQDSENLSGSGSFYEGTD-DK 905  
Qy 1783 VITPLDPTAQLRIMPLPAGG-----PSIS---QGLPA-----SRYNATA 1817  
Db 906 VRRDLATDLSLIEVKLAAGRAVGRKDEFTAEKEATPPTSADKSGLSREFDHRKANDKLDTV 965  
Qy 1818 ADALAALVDAASAAPQ-----MDVSKTKE-----SKH 1844  
Db 966 LEKSEHIDSKEHAKSEMGGKVELFGIGITYDOASTKELITTKDTSPEKTERGLSSVP 1025  
Qy 1845 EAARLEENLSR-----SAAVSEQQOLE-----QKTLEVEKRSVOCYL 1882  
Db 1026 EVAEVEPTTKADQGLDFAATKAEPQSLDIKYDSFGOMASGMNVDAGKAIELKFVFAQELT 1085  
Qy 1883 TSSAPPSKGPQPHSVVYSEAG--KDKGPPPKSRYEELTRGKTTTAAFNDFVITRQ 1940  
Db 1086 LSSE-----APOEADSFMGVESGHKEGKVNETEVKEK-----VTKPLDV-----HQ 1128  
Qy 1941 IASDKDARERGSSQSSSLSSHRYETPPSDAIEVSPASS-----PAPP- 1985  
Db 1129 EAVDK---EESYESGHEHSLTMESLK-PDEGKETSPETSILQIEVALKLSVEIPCPPP 1184  
Qy 1986 --QEKLOT-----YQPEVVKANQANDPTROYEG-PLHHVYRPOOES-----PSPQQQ 2029  
Db 1185 VSEADLSDEKGEVQMEFIQLPKEESTETPDIPALPSDVTPQPEAIVSEPAEVPSEEEE 1244  
Qy 2030 LPPSSQEGMGQVPTRLITLADHICQITQDFARNQVSSQTQQQPTTSTFQNSALY 2089

Db 1245 IEAGEYD-----KLLFRSDTL-----QISDLLVSESEREFVETCPGELK 1284  
Qy 2090 S-----TPVTKT-----SNRYSPEQAQSVHHQRPQS-----RVSPEN 2123  
Db 1285 GVSVSVTIEDFTVVTQTTDEGESGSHSVFAAPAQPEERPRPHDELEIEMAAEA 1344  
Qy 2124 LVDSKRSRPG--KSPERSHVSSEPIEP-----ISPPQVYVHEKQDSLL 2166  
Db 1345 QAEPKDGSPDAPATPEKEEVAFSEVKTETYDDYKDETTIDDSINDADSLWVDQTDDDRSI 1404  
Qy 2167 LLSORGAEPAQR--NDARSP-----GSIS-----YLPSTFTKLENT 2201  
Db 1405 LTEOLETTPKKEAEKDAARRPSLEKHKRPKFTGCRGRISTPERKVAKEPSTVSRDEV 1464  
Qy 2202 SPVMSKKQETFRKLNSGGGSDMAAAQPCQTEIFNLPAVTTSGSVSRGHSFADPASNL 2261  
Db 1465 RKKAVYKKAELAKK-----SEVQAHSPSRKLLILPAI-----KYTRPTH-- 1503  
Qy 2262 GLEDIIIRKALMGSEF-----DKVEDHGVVMS-----QPMGVVPGTANTSV 2301  
Db 1504 -LSCVKKRKTAAAGDLAQAPGAFKQAKDKVTD-GISKSPKRSSLPPRESSILPPRG-- 1558  
Qy 2302 VTSGETREEDGDPHSGGVCKPKLISKNSRKSIPGOGYLGCTERPSSVSY----H 2357  
Db 1559 -VSGD--REENSFSLNS-----ISSARTRRSEPIRRAGKSGTSTPTTTPGSTAITPG 1608  
Qy 2358 SEG DYHRTQCGAWEDRPSSTGTFPNPLTMR-----MLSSSTPTPTIACAPS 2406  
Db 1609 TPPSYSRKTPG-----TPGTPSYRTPGTPKSGILVPSEKKVAIIRTPPKSPATPKQ 1660  
Qy 2407 --AVNQAAPHOON 2417  
Db 1661 LRLINQPLPDLKN 1673

RESULT 39  
NFM\_MOUSE  
ID NFM\_MOUSE STANDARD; PRT; 848 AA.  
AC P08553: Q61961;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)  
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).  
GN NEFM OR NFM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87246694; PubMed-3036526;  
RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;  
RT "Structure and evolutionary origin of the gene encoding mouse NF-M,  
RT the middle-molecular-mass neurofilament protein.";  
RL Eur. J. Biochem. 166:71-77(1987).  
RN [2]  
RP SEQUENCE OF 322-540 FROM N.A.  
RX MEDLINE-87158637; PubMed-3103856;  
RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosveld F.;  
RT "Cloning and developmental expression of the murine neurofilament  
RT gene family.";  
RL Brain Res. 387:243-250(1986).  
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND











Db 1381 TSSRTVESGLEVAAPSVTPQAGTALLAPFPPTQRCVSNPPCETHETGTTTATTVTNNM 1440  
QY 1618 ITSQOMQVNLPRDVARGLSPREPLG-----LPYPATRGILDTNPPPIIL 1663  
Db 1441 SNQD-----PPPAASDOGEVESTQSDVNTSSAITVTSSTUTRAVTVTQSTP--- 1492  
QY 1664 VPHPGGTSTPPMDRITYIPGTQITPPPPYNSASMPGHPHAAASAEREREREKE 1723  
Db 1493 VP--GPSVPPEELQVSPGPRQLPPRQL-----LQASASTALMGESAEVLSA 1537  
QY 1724 RERERIAA---SSDLYLRPSEQRCRGSHGVRSPPSVRTQTML---QORSVFQ 1777  
Db 1538 SQTPELPAVDLSGTSPSSQESAGSAVATVVVQPPPTQSEVDQLSLPQELMAEAQA 1597  
QY 1778 TNGTSVITPLDPTAQLRMPLPAGGPSISQGLPASRYNTAADAALVDAASA-----P 1832  
Db 1598 GTTILMVTGLPEE-----LAVTAAEAQA---AATEAQAIAIOVLOAAQOAVMGTE 1651  
QY 1833 QNDVSKT-----KSKHEAARLENLRSAAVSEQQOLEQKTLVEKRSVQCLYTSSA 1886  
Db 1652 PMDTSEAAATVTAELGHLAEGQGGATTPIVLTQQLAALVQOQLOEAQAQOQH 1711  
QY 1887 PSGRQPHSSVYSEAKDKGPPKSRYEELRGTGTTTAAFNIDVITRQTASDKD 1946  
Db 1712 LPTALAPADSL-----NDPAIESNCLNELAGTVPSTVAL----- 1746  
QY 1947 ARERGSQSDSSSLSHRYETPDAIEVISPASSAPPQ-----EKLQTYQPEVVKA 1999  
Db 1747 -----LPSTATESLAPSNFTFVAPQPVVAPAKLQAAATLVEA 1785  
QY 2000 QAENDPTROYEGPLHYRPOQESPPOQ-----LPSSQAEGMG 2040  
Db 1786 NGIESLGKVDLP-----PPSKAPMKENQWFDGVIKGTVMVTHYFLPDD----- 1834  
QY 2041 QVPRHRLITLADHICQITQDFARNQVSSQTPOQPPTSTQNSALVSPVPTKTSNR 2100  
Db 1835 AVPSDDDLGTVPDY-----NOLKKQELQ-----PGTAKFRVAGINACA 1873  
QY 2101 YSPESQASVHQHPRGSRVSPENLVKSGRPGKSPERSHVSSEPYEPISPPQVPVYHE 2160  
Db 1874 RGFSEISAFKTCLPFGFAPCAI-----KISKSPGAHLT---WEPSPVTSKIIIEY 1923  
QY 2161 KODSLLLSQRGAEPAEQNRDARSPTSISYL-----PSFF---TKLENTSPMYKSKQ 2210  
Db 1924 SVYLAIQSSQAGG-----LKSSTPAQLAFMRVYCGPSPCLVQSSSLSNAHIDYTRPA 1978  
QY 2211 EIFR--KLNSGGGSDMAAQPGETE 2235  
Db 1979 IIFRIAARNEKYG-----PATQV 1997

RESULT 41  
ZEP1\_HUMAN STANDARD: PRT; 2717 AA.  
AC P15822;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ZINC FINGER PROTEIN 40 (HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-BINDING PROTEIN 1) (HIV-EPI) (MAJOR HISTOCOMPATIBILITY COMPLEX BINDING PROTEIN 1) (MBP-1) (POSITIVE REGULATORY DOMAIN II BINDING FACTOR 1) (PRDI1-BF1).  
GN HIVP1 OR ZNF40.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90169514; PubMed-2106471;  
RA Fan C.M.; Maniatis T.;  
RT "A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence.";

Genes Dev. 4:29-42(1990).  
[2]  
RN STRUCTURE BY NMR OF 2113-2142.  
RP MEDLINE-91064333; PubMed-2248949;  
RX Omichinski J.G.; Clore G.M.; Appella E.; Sakaguchi K.; Gronenborn A.M.;  
RA "High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution.";  
RT Biochemistry 29:9324-9334(1990).  
[3]  
RN STRUCTURE BY NMR OF 2087-2142.  
RP MEDLINE-92232684; PubMed-1567844;  
RX Omichinski J.G.; Clore G.M.; Robien M.; Sakaguchi K.; Appella E.; Gronenborn A.M.;  
RA "High-resolution solution structure of the double Cys2His2 zinc finger from the human enhancer binding protein MBP-1.";  
RT Biochemistry 31:3907-3917(1992).  
CC -|- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1. IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT IN T-CELL ACTIVATION.  
CC -|- SUBCELLULAR LOCATION: NUCLEAR.  
CC -|- INDUCTION: BY MITOGEN AND PHORBOL ESTER.  
CC -|- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH ZINC-FINGER IN-BETWEEN.  
CC -|- SIMILARITY: STRONG, TO HIVP2.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; X51435; CAA35798.1; -;  
DR PIR; A34203; A34203.  
DR PDB; 3ZNF; 15-JAN-92.  
DR PDB; 4ZNF; 15-JAN-92.  
DR PDB; 1BBO; 31-OCT-93.  
DR TRANSFAC; T00497; -;  
DR MIM; 194540; -;  
DR InterPro; IPR000822; -;  
DR Pfam; PF00096; zf-C2H2; 5.  
DR PRINTS; PR00048; ZINC-FINGER.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 4.  
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
KW Nuclear protein; Repeat; 3D-structure.  
FT DOMAIN 406 456 ZINC FINGERS.  
FT ZN\_FING 406 428 C2H2-TYPE.  
FT ZN\_FING 434 456 C2H2-TYPE.  
FT DOMAIN 803 806 POLY-SER.  
FT ZN\_FING 958 981 C2HC-TYPE (POTENTIAL).  
FT DOMAIN 2087 2139 ZINC FINGERS.  
FT ZN\_FING 2087 2109 C2H2-TYPE.  
FT ZN\_FING 2115 2139 C2H2-TYPE.  
FT STRAND 2088 2088  
FT TURN 2090 2092  
FT STRAND 2095 2095  
FT TURN 2099 2108  
FT HELIX 2109 2109  
FT TURN 2115 2116  
FT STRAND 2123 2124  
FT STRAND 2127 2135  
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;  
-----  
Query Match 2.3%; Score 295; DB 1; Length 2717;















Qy	1222	YDN1KNAREGT--RSPTAHEISLKRSYESVEGNIKQOMSMBRESVPSPAPLEGLICRALPRG	1280
Db	793	KDFYKN--GTVNAPOLPEMDL-----AGTRSLASVSADAЕVARKSVY--	835
Qy	1281	SPHSDLKERNVLSCS1WQGTPTRATTFESFDGLKYPKQIKRESPTPRAPEGAITKGPYDG	1340
Db	836	-----SETVV-----EDSRTGL-----PPVTDENHVIVKTD-----	861
Qy	1341	ITTKEMGRSI---HEIPQDILTOESRKTPPEVQSTPRIIEGSIISOGTPIKFDNNSGQS	1397
Db	862	-SQLEDGYCVFNKYTV-----LPSPVQDS-----ENLSGES	893
Qy	1398	AIKHNVKSLITGFSKLSRGM-PPELTVPENIKVVERKVEDVKAGETVRSRHHISVYSSGP	1456
Db	894	GTFYEGTD-----DKVRRDLATDLSLI---EVKLAAGARVKDEFVSDKEASAHSGDKSGSL	946
Qy	1457	S-----VLRSITHEAPK--AQLSPGIYDDTSARRTPVSYQNTMSRSGSPM	1498
Db	947	SKEFDQEKKANDRLDVLEKSEBEHADSKEHAKKTEAGDEIETGLGVYEQALAK----	1002
Qy	1499	MNRTSDVTIPPKNSTNHEKSKTLTTPQRESIPAKSPVPGVDPVVSHSPFDPHHRGSGPAGE	1558
Db	1003	-----DLISPTDASSKEAKGL-----SSVPEI-----	1025
Qy	1559	VYNSHLPTQLDPAAMPHRALDPAANAAYLFQRLSPTPGYPQOYLAYMENTRQTLNDYI	1618
Db	1026	-----AEVPSKKVEQGLD-----FAVQQLDVKISDF-	1053
Qy	1619	TSOOMQVNLRPDVARGLSPREQPLGPYPATRGIIIDLTNMPTTILVPHPG----	1674
Db	1054	--GOMASGLNIDRRATE-----LKLEATQ-----DMTPSKKAQЕADAPMGVESH	1098
Qy	1675	MDRITYIPGQITPPPPPNYSMSFGHPHTHAAASAERERERERERERERERERERERERER	1734
Db	1099	MKEGTKVSETV-----KOKVAKPOLVHQEAV-----DKESYESSGEHSLATMES-	1144
Qy	1735	DLVLRGSEQOPGRGS-----HGYVRSP-SPSVRTQETWLOORSV-----FOGTNGT	1781
Db	1145	-LKADEGKETSPESSLIQDEAVKLSVEICPPAVSBAADLATERADVOMEFIQGPKEE	1203
Qy	1782	S-----VITPLDPTAQLR--IMPLPAGGPSISQGLPA--SRYNT-----AADALAALVDAAA	1829
Db	1204	SKETPDISITPSDVAEPLHETIVSEPAEQSEEEIEAQGEYDKLLFRSDTL-QITDGLV	1262
Qy	1830	SAPQMDVSKTKSKHB-----AARLEENLRSRSAVSBQOQ-----LQOKTLEVE	1874
Db	1263	SGAREEFVETPSEHKGVIESVVTIEDDFITVQVTTTDEGESGSHSVRFAALEQP--EVE	1320
Qy	1875	KRSVQCLYTSSAPPSGKOPHSSVYV--SEAGDKGPP-----PKSRVEEELRTRGKT	1925
Db	1321	RR-----PSPHDEEFVEEAAEAQAEPKOGSPAPASPREEVALSEYK	1365
Qy	1926	TITAAINFIDVITRQTASDKARSGSSDSS-----SSLSH	1964
Db	1366	TETVDYKDETT1DDSIMDADSLWVDTOQDDRSIMTEQLETIPKEAKAEARRSSLEKH	1425
Qy	1965	RYETPSDAIEVISPASSAPPOEKIQTQPEVVKANQENDPTQYGGPL-HHYRPOQES	2023
Db	1426	RKEKPFK-----TGRGRISTPFRKAKPESTVSRDEYRRKKAYVYKAEALAKTEVQAHS	1480
Qy	2024	PSQOQLPPSSQAEGMGQVPRTHLITLADHICQIITQDF--ARNOVSSOTPOOPTSITF	2081
Db	1481	PSKRFILKPAKYTRPTHUSCVKRTTAAGESALAPSVFQAKDKVSDGVTKSPKRRSS	1540
Qy	2082	QNSFSAIVSTPVYRKTTS-----NRYSPESQAQSVHHORPGSRVSPENLVDKSRGS-----	2131
Db	1541	LPRPSSIL--PPRRGVSGDRDENSFLNASSISS--SARRTTRSEPIRRACKSGTSTPTTP	1596
Qy	2132	-----RPGKSPERSHVSPEYPISP--POQVYVHEHKODSLLLLSORGAEPABQ-----R	2179
Db	1597	GSTAITPGTGP--SYSSRTPTGPGPSPYPRTPHTGTPKSAIIV-----PSBKVAILIR	1648

```

Qy 2180 NDARSPGSIYLPSPFFTKLENTSPMVSKSKQEIIFRKLNSSGGSDMAAAGP----- 2232
      :|| :      :|| :      :|| :      :|| :      :|| :
Db 1649 TPKKSPAT-----PKQLRLINQPLDLKNVSKI-----GSTDNIKYQPKGGQVQIV 1695
      :|| :      :|| :      :|| :      :|| :      :|| :
Qy 2233 TEINLPAVTTIS-GSVSRGHSFADPNASNLGLEDIIR-----TALMGSFDDKVEDHGVV 2285
      :|| :      :|| :      :|| :      :|| :      :|| :
Db 1696 TKKIDLSHVTSKCGSLKNIRH--RPGGRVKIESVKLDFKEKVAQKVGSLDN--AAH--- 1748
      :|| :      :|| :      :|| :      :|| :      :|| :
Qy 2286 MSQPMGVVPGTGANTSVVTSGETRREEGDPSPHSGVGCKPKLISNSRSKSKSPIPGQCYL 2345
      :|| :      :|| :      :|| :      :|| :      :|| :
Db 1749 -----VPGGNGVKIDSQKLNREHAKARYDHGA---EIIQSPGRSS----- 1787
      :|| :      :|| :      :|| :      :|| :      :|| :
Qy 2346 GTERPSSVSVHSEG 2360
      :|| :      :|| :      :|| :      :|| :      :|| :
Db 1788 -VASPRRLSNVSSSG 1801
      :|| :      :|| :      :|| :      :|| :      :|| :

RESULT 44
MRSP_STAAU
ID MRSP_STAAU STANDARD; PRT: 1637 AA.
AC P80544; Q9ZFG2;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METHICILLIN-RESISTANT SURFACE PROTEIN PRECURSOR.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID:1280;
[]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE 1061.
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "pls, a large repeat-rich surface protein of methicillin resistant
RL Staphylococcus aureus";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
[]
RP PARTIAL SEQUENCE.
RC STRAIN-ISOLATE 1061;
RX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelae J., Vuento M., Kuusela P.;
RT "Purification and characterisation of a plasmin-sensitive surface
RL protein of Staphylococcus aureus";
RL Eur. J. Biochem. 236:904-910(1996).
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF115379; AAD09131.1; -.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
KW Signal.
FT CHAIN 1 48 . POTENTIAL.
FT DOMAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1637 AA; 175473 MW; 75BE9ADB469BD309 CRC64;

```

```
Query Match          2.3% Score 289; DB 1; Length 1637;
Best Local Similarity 19.2%; Pred. No. 0.00066;
Matches 263; Conservative 170; Mismatches 460; Indels 476; Gaps 60;

QY 503 QQIARPSQEKEVEEKEDKAEKTEKKBEKK-----DEEEKDEKDSKE--NTKKKRIDG 556
```



Db 74 EQSSANSNEDSPEQVDVTKDTTEQASSTEEKKANTTEQASTEEKADTTQOATEAPKAG 133  
Qy 557 TAE-ETEAREQATPRGRKTANSQGRKGRITRSMTEAAASAAAAATEEPPLPPPP 615  
Db 134 TDKVETEE---APRAEETDKA---TTEAPKAEETDKATEAP----- 170  
Qy 616 EPISPEVETSWTBEEMEVAKGLVHGHRNWAIAKMWGTKEAOCKNFYNYKRRNL 675  
Db 171 ---KTE---ETDKATTEAPAA----- 186  
Qy 676 DNLLQHQHOKTSRKPREEDVYQCESVASTYSAQDEDEIEASNEENPEDSEVEAVKPS 735  
Db 187 -----EETSKAATEAPKAEETSKAATEAPKAEETKATATEAPKTEETDKVETEE 238  
Qy 736 DSPENATSRGNETAPEVELEPT---TETAPSTSPS---LAVPSTKPAED----- 777  
Db 239 APKAEETSKAATEAPKAEETKVEAPAEETKAEETKATATEPATEDTNAKSNSNAQPS 298  
Qy 778 ESVETQVNDISAETAQMDVDOQBSH-----AEGSV 810  
Db 299 ETERTQVVDVTKADLYKKSEVTEAEKAEIEKVLPKDISNLSNEEIKKIALSEVLKETANK 358  
Qy 811 CDPPATKADSDVVEVRPE-NHAS-----KVEGDNTKERDLDRASEKVEPR 856  
Db 359 ENAQPRATFRSVSNARTNNVSATLRAAAQDPTVTKGTCGNTAHDGDIHKTYKEFP 418  
Qy 857 DEDLVAAQINAAQPE-----POSNDSSATCSADEVDGEP 893  
Db 419 NEGTLTANTNF-NENTGCKGALENDKIDNKDFTIIVPANNQGTTCAD----- 470  
Qy 894 ERQRMFMDSPKSLNPTGTSILVSSPL-----KPNPLDLPLQLOHRAAV 936  
Db 471 GWGFMFTQGGQDFLN-QGGILDRGDMANASGFKIDTAYNNVNGKVDKLDADKTNLS- 527  
Qy 937 IPPMVSCPTPCNIPCTIPVSGYALYORH-----IKAMHESAL 972  
Db 528 -----QIGAAGVGYGTFVKNGADGVTNOVGQNALTKDKPVNKIYIADNTNH 575  
Qy 973 LEEQRQEQIDL-----ECRSSTSPCTSKSPNREWEVLOPAPHO- 1013  
Db 576 LDGQPHGRLNDVNLDAANTSTIATYAGTKWATTDLDIGKS--QKYNFLITSSHQ 633  
Qy 1014 -----LITNLPGEVRLPTRP-----TRPP-----PPLIPSS----- 1040  
Db 634 NRYSGIMRTNL-EGVTTTTPQADLDDVEVTQKPIPKHTIREFPTLEPGSPDIVQKG 692  
Qy 1041 ---KTTVASEKPSFIMGSSISQGTGTYLTSNHOASTQETPK----- 1080  
Db 693 EDGEKTTTTPKVDPTGDVVERGEPTTEVTKNPVDIEIVHFTPEVPGHKDEFDPNLP 752  
Qy 1081 -----PSVGSISLGLPROQE-SAKSATLPYIKQEEFSPRSNS 1117  
Db 753 DGETEVPKPGKKNPETGEVTPPVDDVTKHGPRAGEPEVTKEEIPFKKREFNPLKPG 812  
Qy 1118 QPEGLLVRAQHEGVVGTAGIAQEGSTRGPTSKISVESTPSLRG-SITQGTALPQTG 1176  
Db 813 EE---KVTQEG-----QTEKTTTPT-----TINPLTGEKVGEGEPTTEVTK 852  
Qy 1177 IPTALVKGSI SRMP-----IEDSSPEKGREAAASKGHVIEKSGHLS---YDNKN 1227  
Db 853 EPVDEITQGGEEVPGQHKDEFNPLDGTGEVPGKPG-IKNPETGEVTPPVDDVTKH 911  
Qy 1228 AREGTRSPRTAHEI-----SLKRSYSEVGNKQGMRESVPSA-PLGGLICRA 1276  
Db 912 GPKAGEPEVTKEEIPFKKREFNPLKPGEEKVTEGQTEKTTTPTINPLTG---EK 968  
Qy 1277 LPRGSPHSD-----LKERTVLUG-SIMQ-----TPRATTESFED-----GLKYPKOIKR 1320  
Db 969 VGEGETTEVTKEPVEITQGGEEVPGQHKDEFNPLDGTGEVPGKPGKKNPETGEV 1028  
Qy 1321 ESPPIRAFGAIGTKGPKVDGTTTIKEMGRSITHEIPRODILTOESRK-TPEVVQSTRPIE 1379  
Db 1029 VTPPV---DDVTKHGKAGEP-----EVTKEEIPFKKREFNPLKPGEEKVTK 1074

Qy 1380 ---GSIQGTPIKFDNNNSGOSATKHNVKSLITGSPSKLSRGMPLLEIVPENIKVVVERGKY 1435  
Db 1075 EGQGTGEKTTTTPTTINPLTGE-----KVGEGETTEVTKEPVEITQFGG 1119  
Qy 1436 EDVKAGET-----VRSRHTSVSSGSPSVLSTLHEAPKAOLSPGI 1475  
Db 1120 EEPQGHKDEEDPNLPIDGTTEVPKPGIKKNPETGEVTPPVDDVTKH-GPKAG-EPEV 1176  
Qy 1476 YDDTSARTPTVSYQNTMSRSGSPMNRSTSDVTIPPKNKSTNHERKSTLTPTQRESIPAKSPV 1535  
Db 1177 TKEEIPYETKRVLDPTMEPGSP--DKVA-----QKGNGE-KTTTTPTTINPLTGEKVG 1227  
Qy 1536 PG-----VDPVVSHP-FDPHHRGSTAGEVYVWSHLPTQDLPAMP 1573  
Db 1228 EGEPTTEVTKEPIDEIVNVAEIPH---GTREEI-----DPNLP 1264

RESULT 45  
CDP\_HUMAN  
ID CDP\_HUMAN STANDARD; PRT; 1505 AA.  
AC P39880; 1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CCAAT DISPLACEMENT PROTEIN (CDP).  
GN CUTL1  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Umbilical vein;  
RX MEDLINE=93250973; PubMed=1301999;  
RA Neufeld E.J., Skalknik D.G., Lievens P.M.-J., Orkin S.H.;  
RT "Human CCAAT displacement protein is homologous to the Drosophila  
homeoprotein, cut1".  
RL Nat. Genet. 1:50-55(1992).  
CC -!- FUNCTION: PROBABLY HAS A BROAD ROLE IN MAMMALIAN DEVELOPMENT AS A  
REPRESSOR OF DEVELOPMENTALLY REGULATED GENE EXPRESSION. MAY ACT BY  
PREVENTING BINDING OF POSITIVELY-ACTIVATING CCAAT FACTORS TO  
PROMOTERS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF TRANSCRIPTS ARE PRODUCED BY  
ALTERNATIVE SPLICING.  
CC -!- DOMAIN: ASN AT POSITION 47 OF THE HOMEOBOX MAY PARTICIPATE IN  
REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND  
HETERODIMERIZATION.  
CC -!- SIMILARITY: CONTAINS THREE CUT-REPEATS AND A HOMEOBOX DOMAIN  
SIMILAR TO THOSE OF DROSOPHILA CUT PROTEIN. CDP MAY BE THE  
MAMMALIAN COUNTERPART OF CUT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M74099; -; NOT\_ANNOTATED\_CDS.  
DR HSSP; P10037; 1A07.  
DR TRANSFAC; T00100; -.  
DR MIM; 116896; -.  
DR InterPro; IPR001356; -.  
DR Pfam; PF00046; homeobox; 1.  
DR PROSITE; PS00027; HOMEOBOX\_1; 1.  
DR PROSITE; PS00071; HOMEOBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein; Repeat;  
KW Repressor; Coiled coil; Alternative splicing  
FT DOMAIN 7 363 COILED COIL (POTENTIAL).  
FT REPEAT 551 610 'CUT'-REPEAT.







Db 1426 AAPGEG-----PAAPTSA PPPNSSSSAPRRPSSLQSL 1459

Search completed: September 8, 2001, 14:45:03  
Job time: 330 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:36:58 ; Search time 28.24 Seconds  
(without alignments)  
6581.654 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPNQGAFSTEQSR.....EREPAPLLSAOYETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11536.5	91.2	2453	2 S60254	nuclear receptor c
2	5395	42.7	1047	2 T46489	hypothetical prote
3	1879	14.9	1495	2 S60255	transcription co-r
4	584.5	4.6	5327	2 T13564	microtubule-associ
5	487.5	3.9	2364	2 A56577	microtubule-associ
6	465.5	3.7	2774	2 A43359	microtubule-associ
7	447	3.5	2464	1 QRM3P1	microtubule-associ
8	444	3.5	7962	2 I38346	elastic titin - hu
9	440	3.5	5170	2 T15348	hypothetical prote
10	437	3.5	2845	2 I49505	adenomatous polyo
11	429	3.4	2649	2 T51023	hypothetical prote
12	415	3.3	2897	2 B48666	cell proliferation
13	411.5	3.3	3924	2 S37431	ankyrin 2, neuropa
14	409.5	3.2	2361	2 T27552	hypothetical prote
15	408	3.2	2187	2 T30826	nascent polypeptid
16	405	3.2	2722	2 T20332	hypothetical prote
17	405	3.2	3942	2 T42730	Bassoon protein -
18	404	3.2	3256	2 A48666	cell proliferation
19	402.5	3.2	1110	2 I51116	NF-180 - sea lamp
20	391.5	3.1	1226	2 T42717	DNA-binding protei
21	389	3.1	1262	2 S15053	hypothetical prote
22	389	3.1	5762	2 A41819	proline-rich pepti
23	381.5	3.0	2215	2 T16871	hypothetical prote
24	380.5	3.0	3938	2 T42761	Bassoon protein -
25	376.5	3.0	2447	2 T16870	hypothetical prote
26	376	3.0	1791	2 T02345	hypothetical prote
27	374.5	3.0	3562	2 A47171	chondroitin sulfat
28	372	2.9	2526	2 T20531	hypothetical prote
29	365	2.9	2843	1 RBHUAP	adenomatous polyo

ALIGNMENTS

RESULT 1

S60254

nuclear receptor co-repressor N-Cor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 08-Oct-1999

C;Accession: S60254

R;Hoerlein, A.J.; Naeae, A.M.; Heinzel, T.; Torchia, J.; Gloss, B.; Kurokawa, R.; Ry

Nature 377, 397-404, 1995

A;Title: Ligand-independent repression by the thyroid hormone receptor mediated by a

A;Reference number: S60254; MUID:96008539

A;Accession: S60254

A;Status: preliminary; nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-2453 <HOE>

A;Cross-references: EMBL:U35312; NID:g1022717; PIDN:AAB17125.1; PID:g1022718

Query Match. 91.2%; Score 11536.5; DB 2; Length 2453;

Best Local Similarity 91.0%; Pred. No. 0;

Matches 2245; Conservative 79; Mismatches 101; Indels 43; Gaps 12;

Qy	1	MSSSGYPNQGAFSTEQSRYPHVSQVYTFPNTRHQQEFAVPDYRSSHLEVSQASQLLQQQ	60
Db	1	MSSSGYPNQGAFSTEQSRYPHVSQVYTFPNTRHQQEFAVPDYRSSHLEVSQASQLLQQQ	60
Qy	61	QQQLRRRPSLLSEFHGSDRQERRTSYEFHFGPSPVDHDSLESKRPRLEQVSDSHFQ	120
Db	61	QQQLRRRPSLLSEFHGSDRQERRTSYEFHFGPSPVDHDSLESKRPRLEQVSDSHFQ	120
Qy	121	RVSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPSPISGQPCGDDQNASPSKLSKEE	180
Db	121	RVSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPSPISGQPCGDDQNASPSKLSKEE	180
Qy	181	LIQSMRDVREIAKVEQOILKLKKQQOOLEEAAKPPPEKPPVSPVVEQKHRSIVQIYY	240
Db	181	LIQSMRDVREIAKVEQOILKLKKQQOOLEEAAKPPPEKPPVSPVVEQKHRSIVQIYY	240
Qy	241	DENKKAEEAHKIEGGLPKVPELYNPQSDTKVYHENIKTNQVMRKKLIILFFKRRNAR	300
Db	241	DENKKAEEAHKIEGGLPKVPELYNPQSDTKVYHENIKTNQVMRKKLIILFFKRRNAR	300
Qy	301	KOREQKICQRYDQLEAWKVKVDRIENNRKAKESKREYVEKQFPEIRKQREOQERFQ	360
Db	301	KOREQKICQRYDQLEAWKVKVDRIENNRKAKESKREYVEKQFPEIRKQREOQERFQ	360
Qy	361	RVGQAGLSATIAARSEHSEIIDLSEQENNEKQMLSVIPPMFMDAQRVKFINM	420
Db	361	RVGQAGLSATIAARSEHSEIIDLSEQENNEKQMLSVIPPMFMDAQRVKFINM	420
Qy	421	NGLMEDPMKVYKDRQFMNVWTDHEKEIFKDFIQHPKFNGLIASYLERKSPVDCVLYYL	480
Db	421	NGLMEDPMKVYKDRQFMNVWTDHEKEIFKDFIQHPKFNGLIASYLERKSPVDCVLYYL	480



QY 481 TKKNENYKALVRRYNGKRRGRNQIARPSQEKVEEKEDKAETKEKEEKKDEEKEKD 540  
DB 481 TKKNENYKALVRRYNGKRRGRNQIARPSQEKVEEKEDKAETKEKEEKKDEEKEKD 540  
QY 541 KEDSKENTKDKIDGRAETEEERQATPRGRKKTANSOGRKRGRITRSMWNEAASAAA 600  
DB 541 KEDSKENTKDKIDGRAETEEERQATPRGRKKTANSOGRKRGRITRSMWNEAASAAA 600  
QY 601 AAATEEPPLPPPEPISPEVTSRWTEEMEVAKGLVEHGRNWAIAKMGVTKSEA 660  
DB 601 AAATEEPPLPPPEPISPEVTSRWTEEMEVAKGLVEHGRNWAIAKMGVTKSEA 660  
QY 661 QCKNFYNYKRRHNLNLLQHKOKTSKPREEDVSOCEVASTVSAQEDIEDTEASNEE 720  
DB 661 QCKNFYNYKRRHNLNLLQHKOKTSKPREEDVSOCEVASTVSAQEDIEDTEASNEE 720  
QY 721 ENPEDSE-----VEAVKPSDESPENATSRGNTBPVALEPTTETAPSTS 764  
DB 721 ENPEDSE-----VEAVKPSDESPENATSRGNTBPVALEPTTETAPSTS 764  
QY 765 PSLAVPSTKPAEDSVETQVNDSTSAETAQMDVDQOEHSAEESVCDPPPATKADSDV 824  
DB 765 PSLAVPSTKPAEDSVETQVNDSTSAETAQMDVDQOEHSAEESVCDPPPATKADSDV 824  
QY 825 EVRYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINAOQRPQSDNDSSATCS 884  
DB 825 EVRYPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINAOQRPQSDNDSSATCS 884  
QY 885 ADEDVGEPEQRMPMDSKPSSLNLTGSIILVSSPLKPNPLDLQLQHRAAIVPPMVSC 944  
DB 885 ADEDVGEPEQRMPMDSKPSSLNLTGSIILVSSPLKPNPLDLQLQHRAAIVPPMVSC 944  
QY 945 PCNIPITPVSGYALYQKHAKHESALLEQORQORQOVIDIECRSSSPCSTSKSPNRE 1004  
DB 945 PCNIPITPVSGYALYQKHAKHESALLEQORQORQOVIDIECRSSSPCSTSKSPNRE 1004  
QY 1005 EVLOPAPHOLITNLPPEGVRLPTTRPPPLIPSSKTTVAASEKPSFTMGSGISQGTPT 1064  
DB 1005 EVLOPAPHOLITNLPPEGVRLPTTRPPPLIPSSKTTVAASEKPSFTMGSGISQGTPT 1064  
QY 1065 YLTSHNOASYTOETPKPSVSGISLGLPQOESAKSATLPYIKOEEFSPRSONSOPEGLV 1124  
DB 1065 YLTSHNOASYTOETPKPSVSGISLGLPQOESAKSATLPYIKOEEFSPRSONSOPEGLV 1124  
QY 1125 RAQHEGVVGTAGALOGESITRGTPTKISVESITPSLRGSIQTCTPALPQIGITEALVK 1184  
DB 1125 RAQHEGVVGTAGALOGESITRGTPTKISVESITPSLRGSIQTCTPALPQIGITEALVK 1184  
QY 1185 GSISRMPIEDSSPEKGREAAKSHVIEGKSGHLSYDNKKNAREGTRSPRTAHEISLK 1244  
DB 1185 GSISRMPIEDSSPEKGREAAKSHVIEGKSGHLSYDNKKNAREGTRSPRTAHEISLK 1244  
QY 1245 RSYESVEGNIKQGMRESVPSAPLEGICALRPGSPHSDLKERTVLSGIMQGTPTAT 1304  
DB 1245 RSYESVEGNIKQGMRESVPSAPLEGICALRPGSPHSDLKERTVLSGIMQGTPTAT 1304  
QY 1305 TESPEDEGLKYPKQIKRESPPITRAFEAGITKGPYDGTITTKEMGRSTHETIPRODILQES 1364  
DB 1305 TESPEDEGLKYPKQIKRESPPITRAFEAGITKGPYDGTITTKEMGRSTHETIPRODILQES 1364  
QY 1365 RKTPEVVOSTRPIIEGSIQGTPIKFDNNSGOSAIAKHNVKSLITGPKSLKRGMPLEIIVP 1424  
DB 1365 RKTPEVVOSTRPIIEGSIQGTPIKFDNNSGOSAIAKHNVKSLITGPKSLKRGMPLEIIVP 1424  
QY 1425 ENIKVVERGKVEDYKAGETVRSRHTSVVSSGSPVLSRTLHEAPKAQLSPGIYDDTSARR 1484  
DB 1425 ENIKVVERGKVEDYKAGETVRSRHTSVVSSGSPVLSRTLHEAPKAQLSPGIYDDTSARR 1484  
QY 1485 PVSQNTMSRSGSPMMNRSTSDVTIPPNKSTNHKRSSTLTPTQRESIPAKSPVPGVDVIVSH 1544  
DB 1485 PVSQNTMSRSGSPMMNRSTSDVTIPPNKSTNHKRSSTLTPTQRESIPAKSPVPGVDVIVSH 1544  
QY 1494 PVSQNTISRSGSPMMNRSTSDVS--SSKSASHERKSTLTPTQRESIPAKSPVPGVDVIVSH 1551  
DB 1494 PVSQNTISRSGSPMMNRSTSDVS--SSKSASHERKSTLTPTQRESIPAKSPVPGVDVIVSH 1551

QY 1545 SPFPDHRHGSTAGEVYWSHLPTQLDPAMPFHRALDPAAAYLFRQLSPPTGYSQYQLY 1604  
DB 1545 SPFPDHRHGSTAGEVYWSHLPTQLDPAMPFHRALDPAAAYLFRQLSPPTGYSQYQLY 1604  
QY 1605 AMENTROTILNDYITSOQMQLNRPDVARGLSPREQPLGLPYPATRGLIIDLITNNPPTILV 1664  
DB 1605 AMENTROTILNDYITSOQMQLNRPDVARGLSPREQPLGLPYPATRGLIIDLITNNPPTILV 1664  
QY 1665 PHPGGTSTPPMDRITYIPGTOITFPPRPNYSASMSPGHPTLHAAAASAEERERERER 1722  
DB 1665 PHPGGTSTPPMDRITYIPGTOITFPPRPNYSASMSPGHPTLHAAAASAEERERERER 1722  
QY 1723 -----ERERERIAAASDLYLRPGSEQRPGSHGYVRSPPSVTOETMLQOQPSV 1774  
DB 1723 -----ERERERERERIAAADLYLRPGSEQRPGSHGYVRSPPSVTOETMLQOQPSV 1774  
QY 1775 FQGTNGSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADAALVDAASAPQM 1834  
DB 1775 FQGTNGSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADAALVDAASAPQM 1834  
QY 1835 DVSKTKESKHEAARLEENLRSRAVSEQQOLEQKTLVEKRSVQCLYTSASFSGKQPP 1894  
DB 1835 DVSKTKESKHEAARLEENLRSRAVSEQQOLEQKTLVEKRSVQCLYTSASFSGKQPP 1894  
QY 1895 HSSVYVSEAGKDKGPPPKRSRYEEELTRGKTTITAANFIDVITRQIASDKDARERGQS 1954  
DB 1895 HSSVYVSEAGKDKGPPPKRSRYEEELTRGKTTITAANFIDVITRQIASDKDARERGQS 1954  
QY 1955 SDSSSLSSHRYETPSDAIEVISPASSPAPPEQKLTQYQPEVVKANOENDPTRQYEGPL 2014  
DB 1955 SDSSSLSSHRYETPSDAIEVISPASSPAPPEQKLTQYQPEVVKANOENDPTRQYEGPL 2014  
QY 2015 HHYRPOQESPSPOQ--LPPSSQAEGMGQVPRHRLITLADHICQIITQDFARNVSSQT 2072  
DB 2015 HHYRPOQESPSPOQ--LPPSSQAEGMGQVPRHRLITLADHICQIITQDFARNVSSQT 2072  
QY 2073 PQQPTSTFQNSPALSVPVTRKTSNRYSPESAQSVHHQRPGRSVSPENLVKSRGSR 2132  
DB 2073 PQQPTSTFQNSPALSVPVTRKTSNRYSPESAQSVHHQRPGRSVSPENLVKSRGSR 2132  
QY 2133 PGKSPERSHVSSEYEPISPPQVPVHVKQDSLLSLLSOGAEPAPAEORNDARSPGISYLP 2192  
DB 2133 PGKSPERSHVSSEYEPISPPQVPVHVKQDSLLSLLSOGAEPAPAEORNDARSPGISYLP 2192  
QY 2193 SFFTKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPAVTISGVSRRGH 2252  
DB 2193 SFFTKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPAVTISGVSRRGH 2252  
QY 2207 SFFTKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPAVTISGVSRRSH 2266  
DB 2207 SFFTKLENTSPMVKSKKQEIFRKLNSGGSDMAAQAQGTIEFNLPAVTISGVSRRSH 2266  
QY 2253 SFADPASNLGLIEDIRKALMGSDDKVEDHGVVMSQPMGVVPGTANTSVTSGTTRREG 2312  
DB 2253 SFADPASNLGLIEDIRKALMGSDDKVEDHGVVMSQPMGVVPGTANTSVTSGTTRREG 2312  
QY 2313 DPSPHSGVCKPKLISKNSRKSIPGQYLGTERPSSVSSVHSEGDYHROTPGNAWE 2372  
DB 2313 DPSPHSGVCKPKLISKNSRKSIPGQYLGTERPSSVSSVHSEGDYHROTPGNAWE 2372  
QY 2372 DRPSTGSTQFPYNNPLTWRMLSSPTPTIACAPSAVNOAAPHQONRIWEREPALLSAQY 2432  
DB 2372 DRPSTGSTQFPYNNPLTWRMLSSPTPTIACAPSAVNOAAPHQONRIWEREPALLSAQY 2432  
QY 2433 ETLSDSD 2440  
DB 2433 ETLSDSD 2440  
QY 2446 ETLSDSD 2453  
DB 2446 ETLSDSD 2453

## RESULT 2

T46489  
hypoetical protein DKFZp434M075.1 - human (fragment)

C;Species: Homo sapiens (man)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C;Accession: T46489  
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000



A:Reference number: 223035

A:Accession: T46489

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1047 <AAA>

A:Cross-references: EMBL:AL137641

A:Experimental source: adult testis; clone DKF2p434M075

C:Genetics:

A:Note: DKF2p434M075.1

Query Match 42.7%; Score 5395; DB 2; Length 1047;  
Best Local Similarity 99.6%; Pred. No. 3.3e-202;  
Matches 1043; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1394 SGQSAIKHNKSLITGSKSRGMPPLIEIVPENIKVYERKGYEDVKAGETVRSRHTSVVS 1453
DB 1 SGQSAIKHNKSLITGSKSRGMPPLIEIVPENIKVYERKGYEDVKAGETVRSRHTSVVS 60
QY 1454 SGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSPMNRSTSDVTIPPKNST 1513
DB 61 SGPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRSPMNRSTSDVTISSNKT 120
QY 1514 NHERKSTLTPTQRESIPAKSPVGVDPVSHSPDPHRRGTAAGEVYWSHLPTQLDPAMP 1573
DB 121 NHERKSTLTPTQRESIPAKSPVGVDPVSHSPDPHRRGTAAGEVYWSHLPTQLDPAMP 180
QY 1574 FHRALDPAALAYLFQROLSPTPPGYPOLYAMENRTQTLINDYITSOOMVNLRPDVAR 1633
DB 181 FHRALDPAALAYLFQROLSPTPPGYPOLYAMENRTQTLINDYITSOOMVNLRPDVAR 240
QY 1634 GLSPREQPLGLPYPATRGIIDLTNNPPTILVPHPGGTSTPPMDRITYTPGQITFPFRPY 1693
DB 241 GLSPREQPLGLPYPATRGIIDLTNNPPTILVPHPGGTSTPPMDRITYTPGQITFPFRPY 300
QY 1694 NSAGMSGCHPTHAAAAASAERERERERERERERERERERERERERERERERERERERER 1753
DB 301 NSAGMSGCHPTHAAAAASAERERERERERERERERERERERERERERERERERERERER 360
QY 1754 VRSPSPVQTOETMLQORPSVFQCTNGTSVTITPLDPTAQLRIMPLPAGGPSISQGLPASR 1813
DB 361 VRSPSPVQTOETMLQORPSVFQCTNGTSVTITPLDPTAQLRIMPLPAGGPSISQGLPASR 420
QY 1814 YNTAADALALVDAASAPQMDVSKTESKHEAARLEENLRSAAYSEQQOLEQKLTLEV 1873
DB 421 YNTAADALALVDAASAPQMDVSKTESKHEAARLEENLRSAAYSEQQOLEQKLTLEV 480
QY 1874 EKRSVQCLYTSSAPPSPGKQPPHSSVYVYSEAGKDKGPPPKSRYEELRTRGKTTTAAFI 1933
DB 481 EKRSVQCLYTSSAPPSPGKQPPHSSVYVYSEAGKDKGPPPKSRYEELRTRGKTTTAAFI 540
QY 1934 DVIIITROIASDKDARERGSSQSSSSLSHRYETPDAIEVISPASSPAPQEKLOTYQ 1993
DB 541 DVIIITROIASDKDARERGSSQSSSSLSHRYETPDAIEVISPASSPAPQEKLOTYQ 600
QY 1994 PEVYKNAQENDPTQYEGPLHHYRPOQSPSPQOQLPPSSQAEGMGQVPRTHRLITLAD 2053
DB 601 PEVYKNAQENDPTQYEGPLHHYRPOQSPSPQOQLPPSSQAEGMGQVPRTHRLITLAD 660
QY 2054 HICQIITODFARNQVSSQTPOQPTSTFQNSPSALVTPVTKTSNRYSPESQAQSVHHQ 2113
DB 661 HICQIITODFARNQVSSQTPOQPTSTFQNSPSALVTPVTKTSNRYSPESQAQSVHHQ 720
QY 2114 RPSGRVSPENLVDSRSGRPGKSPERSHVSSEYEPISPPQVPPVHKEQDLSLLLSORGA 2173
DB 721 RPSGRVSPENLVDSRSGRPGKSPERSHVSSEYEPISPPQVPPVHKEQDLSLLLSORGA 780
QY 2174 EPABQRNDARSPGISYLPSPFTKLENTSPMVSKKOEIIFKRLNSGGGSDDMAAOPGT 2233
DB 781 EPABQRNDARSPGISYLPSPFTKLENTSPMVSKKOEIIFKRLNSGGGSDDMAAOPGT 840
QY 2234 EIFNLPAVTTSGSVSRGHSPADPASNLGLEDIIRKALMGSFDDKVEDHGVVMSQPMGVV 2293
DB 1 EIFNLPAVTTSGSVSRGHSPADPASNLGLEDIIRKALMGSFDDKVEDHGVVMSQPMGVV 1020
```

```
DB 841 EIFNLPAVTTSGSVSRGHSPADPASNLGLEDIIRKALMGSFDDKVEDHGVVMSQPMGVV 900
QY 2294 PGTANTSVVTSGETRREREGDPSPHSGGCKPKLLISKSNRSKSPIPGOGYLGTERPSV 2353
DB 901 PGTANTSVVTSGETRREREGDPSPHSGGCKPKLLISKSNRSKSPIPGOGYLGTERPSV 960
QY 2354 SSVHSEGDYHRQTPCWAWEDRPSSTGSTQFPYNPILTMMLSTPTPIACAPSAYNAQAP 2413
DB 961 SSVHSEGDYHRQTPCWAWEDRPSSTGSTQFPYNPILTMMLSTPTPIACAPSAYNAQAP 1020
QY 2414 HQQNRWIEREPAPLLSAOYETLSDD 2440
DB 1021 HQQNRWIEREPAPLLSAOYETLSDD 1047
```

RESULT 3

S60255

transcription co-repressor SMRT - human

C:Species: Homo sapiens (man)

C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 05-Nov-1999

C:Accession: S60255

R:Chen, J.D.; Evans, R.M.

Nature 377, 454-457, 1995

A:Title: A transcriptional co-repressor that interacts with nuclear hormone receptors

A:Reference number: S60255; MUID:96008552

A:Accession: S60255

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1495 <CHE>

A:Cross-references: EMBL:U37146; NID:g1045654; PIDN:AAC50236.1; PID:g1045655

Query Match 14.9%; Score 1879; DB 2; Length 1495;  
Best Local Similarity 35.9%; Pred. No. 1.4e-65;  
Matches 572; Conservative 217; Mismatches 55; Indels 268; Gaps 76;

```
QY 993 PCGTS--KSPNREVEVLQAPAH----QLITNLPEGVRLP-----TTRTRPPPPPLIPSSK 1041
DB 25 PCWTSGLPFPVPPREVAKSPHADPSPAFSYAPGHPPLPLGLHDTARVLPRPPTISNPP 84
QY 1042 TTVASEK--PSFI--MGSIISOGTPTGYTLTSHNQASYTQETPKPSVGSISLGLPQOESAK 1098
DB 85 PLISSAKHPSVLERQIGAISQG-----MSVOLHVYPYSEHAKAP-VGPVTMGLPLPMDPKK 138
QY 1099 SATLPYIKOEFSRSONSQPEGLLV-RAQHEGVVGRGA-GAIOBGSITRGTPPTSKISVE 1156
DB 139 LAPSGVAKQQLSPRGAGPPESLGVPQAQASVLRGTALGVSVPGGSITKGIPTRVPSD 198
QY 1157 SIPLRGSITQCTPALPOTGIPTEALVKGSISRMPIEDSSP--EKGREAAASKGHVYIEG 1214
DB 199 SAITYRGSITHGTA-----DVLYKGTITRIIGEDSPSRLDGRGDSLPKGVHVEG 250
QY 1215 KSHILSYDNTKNA-----REGTRSPRTAHEISL-KRSYSEVGNKQGMMSRSPVAP 1268
DB 251 KGHVLSYEGGMSVYTCSEKDGSRSSGPPHETAAKRTYDMMEGRVGRAS-----SAS 304
QY 1269 LEGLICRALP--RCSPHDLKERIVLSGIMOGTIPRATTESEFDGL-KYPQIKRES--- 1322
DB 305 IEGLMGRAIPPERHSPH-HLKEQHHRIGSITQGIIPRSVYEAQEDYLREAKLLRGTPP 363
QY 1323 --PIPAFEGAI-----TKGKP-YDG-ITTIKEMGRSIIHEIPRODILTOESRKTPEW 1371
DB 364 PPPSRDLTEAYKQALGPLKPAHEGLVATVKEAGHSIIHEIPREL-----RHTELP 418
QY 1372 QSTRPIIEGSIQGTPIKFDNN-SQSAIKHNKSLITGSKSRGMPPLIEIVPENIKV 1430
DB 419 LAPRLKEGSIQGTPLKYDTGASTTGSKKHVDKSLIGSPGRTFPVHPDVMAD-ARAL 477
QY 1431 ERGYEDVKAGETVRSRHTSVVSGPSVLRST---LHAPKAQLSPGIYDTSARTPV 1487
DB 478 ERACYE-----ESLSRPGTASSGGSIARGAPVIVPELGRPROSPLTYEDHGA-----P 527
QY 1488 YQNTMSRSGSPMMNRSDVTIP----PNKSTNHERKSTLTPTQRESIPAKSPVGVDPVVS 1543
```



Db 528 FAGHLPRGSPVMTREPTRLQEGSLSSKASQDRKLTSTPRE----IAKSPHSTVPEHPH 583  
QY 1544 H--SPFDPHRGSTAGEVYWSHLTQDLP--AMPHRALDPAAYLAFQRLSPPTGPSQ 1600  
Db 584 HPISPYELLRGVSGVDYLRSHIPLAFDPTSGIPGLD-AAAAAYLPRHLAPNPTPHL 642  
QY 1601 YQLY-----AMENRQILNDYITSQOMVNL-----RDPVARGLSPREQPLGLP 1645  
Db 643 YPVLIRGYPDTALEN-RQIINDYITSQOMHHNTATAMAQRADMRLKGLSPRESSIALN 701  
QY 1646 YPA-TRGIIDITNMP--TILVPHPGSTPPMDRITYIPGQITFPFPPYNSASMSGPH 1703  
Db 702 YAAQPRGIIDLSQVPHLPVLVPPPTGTPATAMDRLAYLPTAPQPFSSR-HSSSPPLSGPG 760  
QY 1704 THL-----AAAASAEER 1744  
Db 761 THLTKPTTSSERERDRDRER 820  
QY 1745 -----PGRPGSHGVRSPP-SVRTQETMLQORPSVFOGTNGTSTVITPLDPTAQLRIMPL 1798  
Db 821 GGGSSSRPASHAHQHSPISTQDA-LQRPVSLHNTGKGLITAVEPSKPTVLRST 879  
QY 1799 PAGGFSIQGLPASYNTAADA-LAALVDAASAPQMDVSKTKESKHEAARLEENLRSR 1857  
Db 880 STSSPV-----RPAATFPATHCPGLGTLGIVYPTLMPEVLLPK-----EAPRVARPERPRA 931  
QY 1858 AAVSEQOOLEKTLVEKRSVQCLYTSSAFSGKP-----QPHSSVYVSEAGKDKGP--P 1910  
Db 932 D-----TGHAFLAKPPARGLEPASS-----PSKGSPEPLVLP 964  
QY 1911 PKRYEELTRGRGTTITAAFINDIITRIQIASDKARERQSQSDS-----SSLSH 1964  
Db 965 PVSCHATTARTPAKN--LAPHASDPAPPASASDPHREKTQSKPFSIQELESRLGYH 1022  
QY 1965 RYETPSDAIEVISPASSAPPQELQYQVPEVVKANAENDPTQRYEGP-----LH 2015  
Db 1023 GSSYSPEGVPEVSPVSPSLTHDGLPKHLELDKSHLEGELRKPQPKVGLGGEAAHLP 1082  
QY 2016 HYRP-----QOESPSQOQLPPSSQAEQMGVPTRLTLADHICQILTODFARNQVSSQT 2072  
Db 1083 HLRLPSPQSSSPLQTAG-----VKGHORVVTLAGHISEVITQDTRHH----- 1129  
QY 2073 PQ-----PPTSTFQNSALYSTPV---RTKTSNRYSPESQAQSVHHQRPGRSVSPENLVD 2126  
Db 1130 PQQLSAPLAPLYSFGA--SCPVLDLRRPSPDLYLPPD-----HGAP-ARGSPH---- 1177  
QY 2127 KSGSRPKSPERSHV---SSEPEYIPSPQ--VPVVHKEQDSLLLSORGAEPAGQND 2181  
Db 1178 -SEGKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPGHSRANSVYPLLYRDEQTEPSRMG 1236  
QY 2182 ARSPGISYLPSPFTKL-ENTSPMVKSKQRIFRKLSGSGGSDMAAQAOPGTEIFNLP 2240  
Db 1237 SKSPGNTSQPAFFSKLTESANAMVSKKQEIKNKLTNRNEPEYNISQDGTIFNMPA 1296  
QY 2241 VTTSGSVSSRGHSPADPAS-NGLEDIIRKALMGSDFDKVEDHGVMSQPMGV---VPGT 2296  
Db 1297 ITGGLMTRYSAQAOEHASTNMGLEAIRKALMGKYDQWEE-----SPPLSANAENPLN 1350  
QY 2297 ANTSV-----VTSETRRECDSPHSGGVCKPKLISKSNRSKSPICOGYLGTRPS 2351  
Db 1351 ASASLPAAMPITAADGRSDHTLTPSGGG--KAKVSRGSRKAKSPAP--GLASGRDPP 1406  
QY 2352 SVSVHSEGDYHROTP--GNAWEDRPSSTGSTQPPYPLNPLTRM-----LSSPTPTPIACAPS 2406  
Db 1407 SVSVHSEGDYHROTP--GNAWEDRPSSTGSTQPPYPLNPLTRM-----LSSPTPTPIACAPS 2406  
QY 2407 AVNOAAPHQNRHIREPAPLLSAQYETLSDS 2439  
Db 1467 G-PLAGPH---HAWDEEPKPLCQYETLSDSE 1495

RESULT 4

T13564  
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N;Alternate names: hypothetical protein EG:49E4.1  
C;Species: Drosophila melanogaster  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13564  
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A;Reference number: Z17689  
A;Accession: T13564  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-5327 <SPA>  
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C;Genetics:  
A;Cross-references: FlyBase:FBgn0025392  
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A;Note: EG:49E4.1  
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.6%; Score 584.5; DB 2; Length 5327;  
Best Local Similarity 19.5%; Pred. No. 1.1e-14;  
Matches 558; Conservative 412; Mismatches 1059; Indels 837; Gaps 124;

QY 53 ASQLLQQQQQQQLRRRPSLLSEFHGSDRPOERTSYEPFHGP-SPVDH---DSLESKR 108  
Db 750 ATGATQATQKRPISRRRGVS-----PSKRA-----PAGSPVQAKPAADLUK 794  
QY 109 PRLEQ--VSDSHFQVSAALVPLVHPLPEGLRAS--ADAKK----- 145  
Db 795 TRLDKGGTTDSSLYSTPSADEATAAKLQDLTASQELDAEQRELDLKEQEVVRELEA 854  
QY 146 ----DPAFGKKEAPSPISGQPC---GDDQN-----ASP 173  
Db 855 VFSDEMCKRQHQIKAEFLREMPAEGTGDGENDEBEYLIIEKEVEQYTEDSIVEQE 914  
QY 174 SKLSKEELIQSMRDVREIAK-----VEQOILKLLKKQQLLEEAAPPEKPV 224  
Db 915 SMTKEEIQKHQDSQSEKRRKSAEEIEAAIAKVEAAERKARLEGASARQDELD 974  
QY 225 PPPVEQKHSRSTVIQIY-----DENRKKAEAHKIFEGLPKVELPLYNQPSDT 272  
Db 975 VEPOSKIAEVDIATAKDIASRTEQLAKPAEE-----ELSSPTPEEKLSKTSDT 1029  
QY 273 K-----VYHENIKTVQVWRKKLLILFFK 294  
Db 1030 KDDQIGAPVDVLPVNLQESLPEEKFSATIESGATTAPLPEDEIRPLDQI-KEDLVIEEK 1088  
QY 295 RRNHARKQREQIKQRYDQQLMEAWKQVDRITENPRKAKESKREYVEKQFPEIRKQRE 354  
Db 1089 YVKEETKEAAIVVATVQTLPAAAPLAIDTILASATKADKANAE---ALGELPDSGE 1144  
QY 355 QQERFQVQRGAGLSATIAARSEHEISEI-----IDGLSQENNEK 395  
Db 1145 RVLPMKMTFEAQQLLRDVIKTDPDEVADLPVHEEADLGLYEKDSQANAKSISHKEESAK 1204  
QY 396 QMRQLSVIPMMFDAQRRYKFINNGLMEDPMKYKDRQFMVWTDHEKEIFDKFIQH 455  
Db 1205 EEKETD-----DEKENKVGIELG---DEPNKV-----DISHVLLKESVQEV 1243  
QY 456 PKNFGLIASYLERKSVDPCLYLYLTTKKNYKALVRRNYKRRGRNQIARPSQEEKVE 515  
Db 1244 AEKVVVETIETVEKK-----QEEIVE 1263  
QY 516 -----EKEDKAETKEKKEE-EKKDEE-----EKEDKSKENTKEDKIDGTAETEE 563  
Db 1264 ATTVITQENQEDLMEQVKDEEHEQKTESGIITEKAKK-SASTPEEKSTSDITSDDEL 1322  
QY 564 REQATPRGRKTANSQGRR-KGRITRSMTNAAAASAAAAATEPPPPPLPPPPPISTE- 621  
Db 1323 AQLADPTTVPPKSAKREDTGSIESPTTIEEAI--EVEVQAKQBAQKVPVPAPEAIKTEK 1380







RESULT 5  
A56577  
microtubule-associated protein MAP 1B - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 16-Feb-1997  
C:Accession: A56577  
R:Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.  
Eur. J. Cell Biol. 57, 66-74, 1992  
A:Title: Identification of two distinct microtubule binding domains on recombinant rat  
A:Reference number: A56577; MUID:92347374  
A:Accession: A56577  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2364 <ZAU>  
A:Cross-references: GB:X60550  
A:Experimental source: brain  
A:Note: nucleotide sequence not given; conceptual translation not complete  
C:Superfamily: microtubule-associated protein MAP1B

Query Match 3.98; Score 487.5; DB 2; Length 2364;  
Best Local Similarity 19.18; Pred. No. 2.3e-11;  
Matches 494; Conservative 360; Mismatches 946; Indels 793; Gaps 117;

Qy 71 LLSFHPGSDRPOERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFORV---SAAVL 127  
Db 79 LLSTHPAN-----KASLTLCF-----EGDWKNSLDRHNLQDFINIKLSASIL 125  
Qy 128 PLVHPLPGLRASADAKDPAGGKHEAPSPISG-----OPC-----GDD----- 168  
Db 126 PEMEGLSEFTYLSSEVSPS---PFDILEPTSGGFLKSKPCYIPPGGRGDSALFAV 182  
Qy 169 -----QNASPSKLSKEELQSDRVD-----RETAKEVQOI-- 199  
Db 183 NGFNMLINGGSEKSCFWKLIRHLDVRVDSILLTHIGDNLPGINSMQKRLAELESQ 242  
Qy 200 -----LKLKKQK-----OLEEEAKPPPP- 219  
Db 243 STNSNDMMKNLISPLDGVVFLNVPENLKNPKNIMKSTEEACFTLQYLNKLSMKPEPL 302  
Qy 220 -----EKPV-----SPPVQKHRSIVQIIVDNRKKAEEAHKIFEG 256  
Db 303 FRVSGNATEPVILFQKMGVGLKMYLVNPKVSSKEMQYFMQQTGNTKDKAE---LILP 358  
Qy 257 LGPKVELPL--YNPSPDFKVVHENIKTNQVMR-----KKLILFKRR 296  
Db 359 NGQEVDPISYLASVSSLIIVHPNAPAEKILRVLPFGNSTOYNILEGLEKUKHLDFLKQP 418  
Qy 297 NHARKO-----RBQICQRYDQLMQAWKQVDRINNRRKAKESKT-REYIEKQ 345  
Db 419 LATQKDLTGQVSTPPVKQVQLKQAD-----SRESLKPATKPLSSKSVRKESKEE 468  
Qy 346 FPETRKQEQERQVQGRGAGLSATTARSEHSEIIDLGSQENNEKQMRQLSVIPP 405  
Db 469 APEATK-ASQVEKTPKVESKE---KVIYKDKPKGVESKPSVTEKEVPSKEEQS-----P 519  
Qy 406 MWFDAEQRRVFINMGLMEDPMVKYDQPMVYTWDEKEIFKDKFTQHPKNGLIASY 465  
Db 520 VKAEVAERAA-----TESKPVTKDKVVKK----- 544  
Qy 466 LERSVPCVLYVYLLTKKNENYKALVRNRYGKRRGRNQIARPSQEEKVEE--KEEDKAE 523  
Db 545 -EIKTPE-----EKKEPKKEVAKKEDKTPLLKDE--KPKKEAKKEIKKEE 594  
Qy 524 KTEKKEEKDEEKEDEKESKENTKEDKIDGTAETEEEREQATPRGRKYANSQGRKK 583  
Db 595 KKLKKEVKEETPLDKAKKEVKKDEKVEK---KEEKEPKKEIK----- 635  
Qy 584 RITRSMTEAAAAAATAAETPPPLPPPPPEPSTETPSTETSWTER-EMEVAKKGLIVE 642  
Db 636 KJSDIKKSTPLSTDKKPAALK---PKVAKKEETPKKEPIAAGLKDKGKGVYKIK----- 688  
Qy 643 HGRNWAATKAVGTGKSEAQCKNFYFNKRRHNLNLLQOHKQKTSRKPREERDVSOQESV 702

Db 689 EGTKTEAATAVGTAAVAAAGV-----AASGPAKE-----LEAE 723  
Qy 703 ASTVSAQDEDEIASNEEENPEDSEV-EAVKPSDESDPENATSRNTEP-----AVELEPTT 757  
Db 724 RSLMSS--PEDLTQDFBELKAAEIDVAKDIKPOLLEDEDEKLAETEPGEAYVQKETE 781  
Qy 758 ETAPSTSPSLAVPSTK-----PADESVEVTONDSI----- 788  
Db 782 SKGSAESPDEGITTEGECEQTEPELEPVKQGVDDIEKFDEGAGFRESSAGDYEE 841  
Qy 789 --SAETAQMDVDQOEHSAEESVCDPPPAKADSDVDEVPVPHENHASKVGDNTKERDL 846  
Db 842 KAETEEAEEPEDEGEDNVSGSASKHSPTEDEEIAKAAADVHIKKEKRESVAGDDRAEDM 901  
Qy 847 DRASEKVEPRDELVLAQQINQAPEPQSDNDSSATCSADEVDGEPERQRMFMDSKPS 906  
Db 902 DEALEKGE-----AEQSEEEGEEDKAEDAREE-DHEPDKTE--AEDYVMA 945  
Qy 907 LLNPTGSLVSS--PLKPNPLDLPLQHRAAVIPPMVSCPCNLIPIGTVPVGYALYQRH 963  
Db 946 VVDKAAEAGVTEDQYDFLGTAKPGVQS-----PSREPAS----- 981  
Qy 964 IKAMHESALLEEQROQRQIDLECRSTSPCGTSKSPNREWEVLQAPHLQILNPEGVR 1023  
Db 982 --SIHDETL-----PGGSESEATASDEENRE---DOPEEFTATSGVTOSTI 1022  
Qy 1024 LPTRPT-----RPPPLIPSSKTTVASEKPS--FIMGGSISQGTPTGTYLTHSHQASV 1076  
Db 1023 EISSEPTMDMSTPRDVMYTDETNNEETSPSOBFV-----NITKYESSLSQ 1070  
Qy 1077 ETPKPSVGSISLGLP-----RQOESAKSATLP--YIKQEFSPRSQNSQPEGLLV 1124  
Db 1071 EYSPVVASFN-GLSDGSKTDTGDRYNASASISPPSMEEDKFKSALRD-----AY 1124  
Qy 1125 RAQHEGVVGRGTAGIQEISITRGTPTKISVESIP-----SLRGSITQGTAL 1172  
Db 1125 RPEETDVKTAGELDIKDVSDERLSAPKSPSLSPSPSPPIEKTPLGERSVNFSLTPNEIKA 1184  
Qy 1173 POTGIPTALVKGISIRMPED--SSPEKGREAAASKGHVIEKSGHILSYD---NKN 1227  
Db 1185 SAEGEAT-AVYSPGVTAQVVEHCASPEEKTLEVVSPSQSV-TGSAGHTPYQSPDTDEKS 1242  
Qy 1228 AREGTRSPRTAHEISLRSY-ESVEGNIKQGMRESVSPAPLEGLICRALPRGSPHSDL 1286  
Db 1243 SHLPTVENTAQAQVPVSEFTEAKDENERSISPMDEV-----PDSES 1286  
Qy 1287 KERTVLGSSIMOGTPRATES-FEDGLKYPKOI---KRESPIRAFEGAITKGPYDGIT 1342  
Db 1287 PIEKVLSP--LRSPPLIGSESAYEDFLSADDKALGRRSESP---FEGKNGKQGFSDRES 1340  
Qy 1343 TIKEMGRSIEH-----IP-RODILLTOESRKTPEV--QSTRPIIEGSI-SQGTPI 1388  
Db 1341 PVSULTSLDYQDKQEEKRAGFIPKEDEFSPEKKAADAEIMSSQSALALDERKLGDSGPT 1400  
Qy 1389 KFDNNNGSALKHNKSLITGTSKLSRGMPLLEIVPENIKVVERKGYEDVAGETVRSRH 1448  
Db 1401 QVD-VSQFGSKFEDTKMSISEGTVSDKSATP-----VDEGAETYSHMEGVASVS 1449  
Qy 1449 TSVVSSGSPVLRLSTLHEAPKAQLSPGIY-----DQT---SARRTPVSVQNTWSR 1494  
Db 1450 TA-----SVATSSFPPEPTDDVSPSLHAEVGSPHSTEDDLSVSVVQTPTTFOET--- 1500  
Qy 1495 GSPMNRSTDTIIPNKSTNHERKSTLTPTORESIPAKSPVGVDPVVSHPFDPHGRS 1554  
Db 1501 -----EMSPSKE-ECPRPMISIPDFSPKTSKRTSPVQD-----HRSE 1537  
Qy 1555 TAGEVYVSHLPTQLDPAHPFHR-ALDPAATAAAYLFQRLSPPTGYPSS--QYQLYAMENRQ 1611  
Db 1538 Q-----SSMSIEGQESPEHSLAND-----FSRQ---SPDHPTVCAGMLHITENGT 1581  
Qy 1612 TILNDYITSOQMQLNRPDVARGLSPREQPLGLYPATRPATRIIDLTNMPPTILVHPGGTS 1671







Db 731 SYIQETIPGYSETEQIISDEIHDEPDERPAPRFTSTYDLSGPEG-----PGPFASQ 786  
QY 818 KADSV-----DVEVRP-----ENHASKVEGDNTKRDLDRASEKVEP 855  
Db 787 AADSAVASSSKTYGAPETELTYPPNVAAPLAEHVVSSA-----TSITECDKLSSATS 842  
QY 856 RDEDLVVAQIQINARPE-----POSNDNS-----SATCSAED 888  
Db 843 VAEDQSVASLTAPQTEETGKSSLLDTVTSIPSSRTEATQGLDYVPSAGTISPTSSLEED 902  
QY 889 -----VDGEPE-----RORMFPMDSKPSLLNPTGSLIVSSP 919  
Db 903 KGFKSPCCDFSVTGESEKKGETVGRGLSGEKAIVGEKVVVTSEK--LSQYAAVFGAP 960  
QY 920 ---LKNPFLDLPOLQRA-----AVIPPMVSCYPCNIPIGTIPVSGYALYQRHIKAM 967  
Db 961 GHTLPPCEPALGEVEERCLSPDDSTVKMASPP-----PSGPPSAATHPF----- 1004  
QY 968 HESALLEEQROQIDLECKRSSTSPCGTSKSPNREWEVLQAPAHQILITNLPGVRLPTT 1027  
Db 1005 HQSPVEDKSEPRDFQEDSWGETHSP--GVSKEDSEE--QTVKPGPEEGTSE--EGKGPPT 1060  
QY 1028 RPTRRPPP-----LIPSSKTTVASEKPSFIMGSSISQGT--PGTYLTSHQAS 1073  
Db 1061 SPOAQDMPVSIAGGOTGCTIQLLPEODKAIYFETGE--AGSNLGAAGTLPGEVRTSTEAT 1118  
QY 1074 YTOETPKPSVGSISLGLPQOESAK--SATLPYIKQEEFSRPS-----QNSQPEGLL----- 1123  
Db 1119 EPQKDEVLRTDQSL--PEDAESLVLSVSPDTTKQENTPRSCSLKEQOPHKDLMPV 1177  
QY 1124 -----VRAQHEGVVRGTA-----GAIQEGSITRG-----PTSK----- 1152  
Db 1178 SPEDQSLSFSESPKETSILDISKQSLSPESLGLTQFGLNLGKEBGPVMAEDDSCH 1237  
QY 1153 ISVESIPS--LRGSTQGTTPALPQTGITEALVKGSIS----- 1188  
Db 1238 LAPVSIPEPHRATVSPSTDETPACTLPGGFSHLSALSDVRKHSFGEITPGGHPMTSDSS 1297  
QY 1189 -----RMPIEDSSPEKGREEAASKGHVYEGKSGHILSYDNIK 1236  
Db 1298 LTKSPESLSPAMEDLAVEWEGKAPGKEPELASETRQKQGLPEKVA--VVEQDLII 1355  
QY 1227 NAREGT-----RSPTAHEISLKRYSVEGNIKQKGMRESVPSAPLEGILICAL 1277  
Db 1356 HKDGADEENKPGQODKTPQKGRDLDEKDTAAELDKGPEKPKDLREDQD----- 1409  
QY 1278 PRGSPHDLKERTVLSGIMQGT-----PRATTESFEDGLKYPKQIK--RESPPIRAFEGAI 1332  
Db 1410 QRAGPPAEKDKASQRTDLQQTQATEPRDRAQERDSEKDKSLELRDRTP----- 1461  
QY 1333 TKGKPYDGIITIKEMGRSIIHEIPRODILTOESR-----KTPVWQSTRIPIEGSISQGTPI 1388  
Db 1462 -----EKDR-----ILVQEDRAPEHSIPEPTQDTR-----APDKGTDD 1496  
QY 1389 KFDNNSGQSAIKHNV-----KSLITGPKSLRGMPPLEIVPENIKYVVERKGYEDKAGETVR 1445  
Db 1497 K-EQKEASEKEQVLEQKDWALKEG-----ETLDQEARTEAQ--KDETLKEDKT-Q 1545  
QY 1446 SRHTSVSVSGPSVLRLHLHAPKAQLSPGIYDDTSA---RRTPVSYQNTMSRGSPPMNR 1502  
Db 1546 GQKSFVEDKTTTSTKETVLQKSAEKADSVEQQDGALEKTRALGLSESPAEGS----- 1599  
QY 1503 SDVTIPPKNKSTNHERK-----STLPTPTQRESIPAK--SPVPG--VDPVVSXSPDPHHRG--- 1553  
Db 1600 -----KAREQEKYKWEODVVQGHRETSPTRGEPVGGQKEPVPVAMEGKSPQEVRYW 1651  
QY 1554 -----STAGEVYVSHLPTQLDPAHPFHRLDPAALAA--AYLFQRLSPTPGYPSQYLIAM 1606  
Db 1652 RDRDITLIQDAYWREL--SCDRKWFVPHELDQGARPRYCEERESTFLDEGPDEQETIPL 1709  
QY 1607 ENT--RQILINDYITSQOMQVNLRPDVARGLSPREOPLGLP-----YPATRGILDLTNMP-- 1659  
Db 1710 QHTPRSPWTSDFKDFQBLPQKQLEVERWLA--ESPVGLPPEEKDKLTRSPFEIISPPAS 1767

QY 1660 -----PTILVPHPGGISTPPMDRITYIPGTQITFP-----RPNYSASMSG- 1701  
Db 1768 PPEMTGORVPSAQOESPVPDTESTAPMRNETTPSWLAEIIPWVWKDRLPPAPLSPAP 1827  
QY 1702 -----HPTHAAAAAERERE-----REREKERERERI 1729  
Db 1828 APPTAPEPHTPPVPSWGLAEYDVAAVQEGAAELEGGPYSPGLKDYKRAEGEREGG 1887  
QY 1730 AAASSDLPLPGSEOPGRPGSHGYVRSPSVRTOETMLQORPSVFOGTNGTSTVITLDP 1789  
Db 1888 AGAPDSSFSFKVPEAGESLATRDETEPEQREPTYPDERSFYADIYEQMMLTGLGP 1947  
QY 1790 TAOIRIMPLPAGG--PSISQGLPASRYNTAADALAAALVDAASAPQMDVSKTESKHEA 1846  
Db 1948 ACTREPPLGASGDWPHLSTKEAAGCNTSAE-----KETSSPA 1987  
QY 1847 ARLEENURSAAVSEGOQLEQKLEVEKRSVQCXYTSSAPPS--GKQPHSSSVVYS- 1901  
Db 1988 S--PQNLSQDTPAFS-----YASLAGPAVPPRQEPDPPGPNVEPSI 2025  
QY 1902 -----EACKDKCQP-----PKSRYEEELRTRGKTTITAAINFIDVITROI 1941  
Db 2026 TPFAVPPRAPISLKSGLDPLNGSTVSCSPDRTPSPKETGRGHWDGTDNSDL----- 2079  
QY 1942 ASOKDARERSQ-----SSDSSSSLSSSHRY-----ETPSDAI 1973  
Db 2080 --EKAREQPEKETRSPSPHPMPGHSSLLWLPETEAYSSLSDDLHSGVRSPLDFPASAF 2137  
QY 1974 EVISPASSPAPPQKLTQYQEVVKANOQ-----ENDPTRQYEGPLUHY 2017  
Db 2138 GFSS--LQAPAPQ-----LPSPAEPRSAPCGSLAFSGDRALALVPGTPTTRH--DEY 2186  
QY 2018 RPOESESPOOQLP--PSSQAEG---MGQVPRHRLITLADHICQIITQDFARNQVSSOT 2072  
Db 2187 LEVTKAPSLDSSLQPLPSPSPGGLLSNLPR----- 2218  
QY 2073 PQOPPTSTFQNSALVSTPVRTKTSNRYSP-----ESQAQSVHHORPG 2116  
Db 2219 --PASPALSEGSSSEATTPIVSSVAERFPGLAEAQSAEGLSGKESAHSLLDWTPL 2275  
QY 2117 SRV-----SPENLVDKSRGSRPKSPERSHVSS 2144  
Db 2276 SPAPSASLDLAPAPAPAPAPGLPGLDGTLPCTRCECTGELTKKPSFLSPSGDHEAN 2335  
QY 2145 EPYE-PISPQ--VPVVEHKODSLLLLS--ORCAEPAEQORDARSFGSISYLPSTFKLENT 2201  
Db 2336 PGGETSLNPPGFTVATAEKEEAAPHAWERGSWPEGAERSRPD-----TLSSSE 2385  
QY 2202 SPVVKSKKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSVSGSVSRGHSFA--DPASN 2260  
Db 2386 QPLRPGK-----SSGG-----PPCSLSSEVAGPQGCATDPRPH 2419  
QY 2261 LGLEDIIRKALMGSDDKVEDHGVMSQPMGVVPGTANTSVVTSGET-----RREGD 2313  
Db 2420 CG-----ELSPSF-----LNPLP--PPSTDSDLSFEARLAGKGGRRVR 2459  
QY 2314 PPSHSGVC-----KPKLISKNSRKSQKPIPGQ-----GYLGE 2348  
Db 2460 PGA--TGCPCPMADETPTPTSASDSSGSSSDSDVPPETECPESITAEALDDEGDGFLPVD 2518  
QY 2349 RPSVSVSHSEGIDYHROTPGWAWEDRPSSTGCTQFPYNYPLTMRMLSTPTPTIAC----- 2403  
Db 2519 KAGVSGCTHHPRGCHDPPPTPLDPPRS-----PPRDPVCMADPE 2558  
QY 2404 -----APSAVNQAAPHQONRI--WEREPAP 2426  
Db 2559 GLSSESGRVERLREKGRPRRAFCRAKPAFPARLDIRGRSRTP 2603

RESULT 7

QRMSP1

microtubule-associated protein MAP1B - mouse















Qy 61 QQQ--OLRRRPSLLSEFH-----PGSRP-----QERRTSYEPFHP--GPSVPVDHDSLES 106  
Db 835 YPHVVETITTTTINVTNINYYDDNEDNVPSSDPATQHOFQOSETSVHSHSDPSVEESDGEGLGS 894  
Qy 107 KRPLEQVSDSHFORVSAVPLVHPUPBGLRASADA--KDPDAPFGKHEAPSPPI-----S 161  
Db 895 KVLGFAKAGMAGVAGVAAVALA---AVGAKAAYDAFEKDDDEDETSHPSPVPPEYQS 951  
Qy 162 QOPCGDDONAS-----PSKLSKEELIOSMDRVDREIAKVEQQILKLLKKQOQLEEE 212  
Db 952 EYQODSAQSHTDFEHMHPESPTEHEKETEEFDH-----SHPESPVLSEKEREHQVTS 1006  
Qy 213 AA-----KPPEPEK-----PVSPPP-----VEQKHSRIVQII----- 239  
Db 1007 TTTTITVTREYNDPEDEOEHOQHSPAPSSHTAEAPHIVETITTTTITVTREFOEPEEL 1066  
Qy 240 ---YDENKKAEAKHIFEGLPKVELPLYNQPSDTKYVYHENIKTNQVMRKKLILFKRR 296  
Db 1067 EYKOEDNSRKSPPSH-----SQENLVTETTTTTITVTSE-----YYDEP 1104  
Qy 297 NHARKOREQKICQYDQIMEAEKKV--DRIENPRKAKESKTRYYEYKOFPEIRKORE 354  
Db 1105 EHFEQETGKSPAPSSHVES--EQVLSPVADPRH--VMETITTTTITVTROFHODESERS 1162  
Qy 355 QOERFORVQORGAGLSATIASEHEISIDG-----LSEOE--NNEKOMRQLSVIP 404  
Db 1163 DSPNRDVV--EAQSIHSHIESHOQSEKDDDSQSGSPVKSDEPVKHESYKFETSTTE 1220  
Qy 405 PMWFDABORRVKFINMGLME---DPMKYKDRQFMNVTWTHEKEIKFKFIQHPKNF-- 459  
Db 1221 TREFDRPE-----LDSERLSPAQSPMEV-----SPITTEENIVRPSVSK 1261  
Qy 460 -----GLIASYLRKSKVPDCLVYLYLTKKNYKALVRRNYGRRGRNQOAIARPSOEK 513  
Db 1262 SEYSGEHPVSVIETKTTTITVTREFYDDQDE-----HENQTSQEEELRSSIPTEEEE 1313  
Qy 514 VEK--EEDKAETKEK--EKKDEEKDEKEDSKENTKEKDIDGTAETEERQAT 568  
Db 1314 HEGSHLFKETTITTTTITVTREFYDEPENVEELQDPQSPAPSSHVESGHASESPAQQOEI 1373  
Qy 569 PRGRTANSQGRKRGRITRSMTNEAAA-----SAAAAAATEBPPLPPPPPISTEP 622  
Db 1374 PQ-----TREFHSPAQYFHEEYESHVLTEQAPLLTEOQHOPESGDE 1418  
Qy 623 VETSRWTEEMEVAKK--GLVHGGRNMAIA--KMVGTKSEAOCKNFYKRRHRNLDNLLQ 680  
Db 1419 SDGEGLSKVLGFAKAGMAGVAGVAAVALAAGAKAAYDALKDDDEDEQERESILR 1478  
Qy 681 QHKO-----KTSRKPREERDVQCESVASTVSAQDE-----DIEA 716  
Db 1479 QERSIDSPHASEQOIEEHEHFEESPVPSEKHHVTITTTTITVTREFYDEHEPLVSQEI 1538  
Qy 717 SNEENPEDE-----VEAVKPSDESPENAT--SRGNTPEAVE 752  
Db 1539 EGEKSPADSEKSLPHVVETTTTITVTREFDKNDESESPVSEKOEPTTVSREYETAE 1598  
Qy 753 LEPT---TETAPT-----SPSLAVPSTKPA-----EDESVE 781  
Db 1599 DEPEHHYETITTTITVTREVIDDSQEMGDDDEPKQESQVETTTTITWISREYDDNDDETR 1658  
Qy 782 TOVNDISAETAQOMDV-----DOQHSAAEGSVCDPPP-----ATKADSDV 824  
Db 1659 SEAGDSHITETTTTITVTREFPHGQPEETEEDVEELPPKIEEDNVSEYSESSTSVSR 1718  
Qy 825 EVRYPENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQOINAQORPEQSDNDSSATCS 884  
Db 1719 EVRPEPHIETTTTITVTREYHNEPEETYDDQD---AAPISFS--QEHQDDDSQAS-- 1771  
Qy 885 ADEDVDGEPERQPMDSKPSLLNPTGSLVSSPLK-----PNPLDLPLOQH 932  
Db 1772 ---HDQDRSPVESEKSVKHTTTTITVTTRQLYDDEASEIRGESPVATEEHEH 1824  
Qy 933 RAAVTPMVSCTPCNIPITGTPVSGVALYVORHIKAMHESALL-----EQQRQOE 982

Db 1825 -----VSTKSDD-----SEQHVPSVIEITTTTITVTREFYDDQDELQREDH 1866  
Qy 983 IDLECRSTSPCGTSKSPNREWEVLQAPAHQOLITNLPEGVRLPTTRPTR---PPPLIPS 1039  
Db 1867 TQSEERRSSIP--TEETHEHDSHLKET-----TTTTTITVTREFYDEPENVEK 1911  
Qy 1040 SKTTVASEKPSFINGGSIQCTPGTYLTSHNOAS--YTQETPKPSVGSISLGLPQQOESAK 1098  
Db 1912 LQDSQFSLSPS-----SHVESEIYPESP-----VAKQOEIPQ 1944  
Qy 1099 -----SATLPYIKQEFSPRSONSQFGLLVRAQHEGVVGRGTAGATQEGSIITRGTPTS 1151  
Db 1945 TREFHEDSPAAYTHEDEYEHQVTEQAP--LLTEQHQHP-----ESGEESDGEFG 1994  
Qy 1152 KISVESIPSLRGSTQCTPALPQTGIPTEAL--VKGSIISRMPIEDSSPEKGREEAASGHV 1210  
Db 1995 S--KVLGFAKAGMAGVAGVAAAP---VALAAGAKAAYDALKKEDD--EEOEERES--L 2045  
Qy 1211 IYEGKSGHILSYDNIKNAREGTRSPRTAHEISLRSVESVEGNIKOGMSMRSPVSAPLE 1270  
Db 2046 IREER-----SFD-----SPHASESQOIEKEHKEFE-----ESPVPSEKH 2080  
Qy 1271 GLICRALPRGSPHSDLKE--RTVLSSGIMOGTPRATTESFEDGLKYPKQIKRESPPIRAF 1328  
Db 2081 HDQSSALPQESVSQPIEKESRTFNDESEFGVKSDDHYTDDQESLSKPE----- 2129  
Qy 1329 EGATKPKPYDGIITIKEMGRS---IHEIPQDILTOESRKTPEVQOSTRIIE----- 1379  
Db 2130 -----SGEAFSQFTSEKQDRSDPIHS--OKEDISQFONESSPEDVKSQEPHDEKPDLE 2183  
Qy 1380 --GSGISQCTPIKFNNSGQSAIKHNKSLITG--PSKLSRGMPPLEIVPENI----- 1427  
Db 2184 ROGYSYSGSPKSPGGS-----ITGLDEKALSCVQEPEDRPFENFAESHEKTE 2231  
Qy 1428 -----KVVERGKYEDVAGETVRSRHTS--VWSSGSPVLSRTLHEAPKAQLSPIYDDTS 1480  
Db 2232 ATSDENLFESDKY---APASVPSPDESNNRVETITTTTITVTREHFEPE-----DDHS 2280  
Qy 1481 ARRTPVSYQNTMRSGSPMMNRTSDVTIPPKNKSTNHERKSTLTPT--QRESIPAKSPVPCVD 1539  
Db 2281 ---YVSESQYSSGSP-----VPEKSDVRVETITTTTITVTREHFAEDEIPTI-- 2327  
Qy 1540 PVWSHSPDFPHRGSTAGEYVWSHLPTQLDPAMPFHRALDPAAAAAYLFQRLSPTPGYPS 1599  
Db 2328 -----VSSH-----DDQAA-----SPVPSSEED 2345  
Qy 1600 QYQLYAMENTROTIL-----NDYITSQOMQVNLDPDVARGLSPREOPLGLPYPATRGII 1653  
Db 2346 VHEQIQTTTTTTTITVTKEHFPDDEIDSEHMN-----ESDKYASGSPVPSSEEDSS 2394  
Qy 1654 DLTNMPPTILVP---HPGGTSTPPMDRITVPGTQITFPFPPRPNYSASMSGPHTHLAAA 1709  
Db 2395 RVETTTTTTITVTREHFEPEDDHSPVQTOEY-----SASESP-----VPSE 2435  
Qy 1710 ASABR-----ER 1745  
Db 2436 KSERVETITTTTITVTREHFEDEHILGQOESDQIPSEISITSENMD---RETSSSPVQ 2492  
Qy 1746 -----GRPGSHGVRSPPSVRT-----QETMLQORPSVFOGTN 1779  
Db 2493 SNRDEEVLPAIAPYKQPTQYGRVDSHDAPASPAESPSPQIAYKQOESQEAHSLFOFQ 2552  
Qy 1780 GTSVITPLDPTAQL-----RIMPLPAG--GPSISQ----- 1807  
Db 2553 RSSVSHEDSPAAYFHDNEDHSDPSVSDRPLLTQEQHQPESGEESDGEFGSKVLGFA 2612  
Qy 1808 ---GLPASRYNTAADALAALVDAASAPQMDVSKTESKHEAARLEENLRSRSAVSEQ 1864  
Db 2613 KKAGMAGVAGVAAAPVALAAVGAkaa---YDALKKDDDEDENPDQEKLLPKS---PERQ 2665  
Qy 1865 QL-----FQKTLVEKRSVQCLYTSSAFPSGKPPQPHSHSVVYSEAGDKGPPPKSYEEELR 1920



Db 2666 VLVNVPDSSEISE--IELEYTS-----PSPTEKSESCY 2699  
Qy 1921 TRGKTTTAAANFIDVITITQASDKDARERGSQSDSSSSSHRYETPPSDAIEVISPAS 1980  
Db 2700 TETVTTTT-----VTREYLD-----PQSVTRSRSPSEH-----DISEQAPES 2738  
Qy 1981 -----SPAPQOE-KLQYIQBEVVKANAQENDPTRYEGPLHHY----- 2017  
Db 2739 PVEQDPYVVEKTTTIRQYHDEPPQIEBEQIPBEVTVLREVYESP--EGDEPEQHYET 2796  
Qy 2018 -----RPOQESPS----- 2025  
Db 2797 KTTTTIYKEVHVPEEDVOISPVHSETSVSEKQLPADQLDEPVTESTTATVTRERYEE 2856  
Qy 2026 PQQLPPS--SOAEGMGQVPRTHRLITLADHICQI--ITQDFARNQVSSQTPOQPTSTFQN 2083  
Db 2857 PEEVPPSGSEADDESHAPKYMETTTTTTITVREYVESEDEHQQSQVORDESAPSEDS 2916  
Qy 2084 SPSALVSTPVTKTSNRYSPSQASV-----HHQRPGRSVSPE 2122  
Db 2917 VKHVIEKTTTTVTEERYEPEDSHSPVPSEDDVHGFKVKTITTTTTTITVTHEHPEPEDPPSDE 2976  
Qy 2123 NLVDKSGSRPGKSPERSHVSSEYEPISPPQVPVH--EKQDSLILLSORGAEPAEQRND 2181  
Db 2977 HVESERTASGVPSEED--SSRIETTTTTVTRHFELED-----DQEHVSEQYSA 3030  
Qy 2182 ARSPGSISYLP-----FFTKLENTSPMVKSKQEIFRKLNSGGSDSMAAAQPGTE 2234  
Db 3031 SGSP-----VPSEKSERVETITNTVTRHFEHEDDIPITVETSHDDPAASSVPSEED 3085  
Qy 2235 IFNLPAVTTSGVSSRGHSFADPASNIGLEDIIIRKALMGSDDDKVEDHGVVMSQPMGVYP 2294  
Db 3086 VHGOIQTTTTTTTTRHVVPDEIDSGRMDLEKYSSESVPVSEEDSSRVIE----- 3138  
Qy 2295 GTANTSVVTSGETREGDPSHGGVCKPKLISKNSRKSIPGQ----- 2342  
Db 3139 -TTTTTTS-----REHFEEDHSHVSEQYSAAG-----SPVPSEKSERVETITTT 3187  
Qy 2343 -----GYLGTGER--PSSVSVHSEH--DVH--RQTPGWAMEDR----- 2374  
Db 3188 TTVTRHFEDEEDYIPESRTHDGDITDQHVPSQSPVSEEDDHVTHEQIIRDDPIDEQ 3247  
Qy 2375 -----PSSGTSGTFYNNPLTMRLSSTPTPTACAPSANQAAPHQONRI-- 2419  
Db 3248 IVESHEYKSESSIPSEQDSTH-----VIETVMSTPITTSERYDPEVEKDVVESADDEIDS 3301  
Qy 2420 -----WEREPAPLLSAQYETLSDSD 2440  
Db 3302 STAOYKSESPVQTEKSLLLAKQQQSGDESD 3333

RESULT 10  
I49505  
adenomatous polyposis coli protein - mouse  
N:Alternate names: APC  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 13-Aug-1999  
C:Accession: I49505  
R:Su, L.  
Science 256, 668-670, 1992  
A:Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the  
A:Reference number: I49505; MUID:92263101  
A:Accession: I49505  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2845 <RES>  
A:Cross-references: GB:M88127; NID:g191991; PIDN:AAB59632.1; PID:g191992  
C:Superfamily: adenomatous polyposis coli protein

Query Match 3.58; Score 437; DB 2; Length 2845;  
Best Local Similarity 19.4%; Pred. No. 2.7e-09;  
Matches 423; Conservative 271; Mismatches 831; Indels 658; Gaps 91;

Qy 347 PEIRKQEQOE--RFQVRGQRGAGLSATIARSEH-----EISEIIDGL 387  
Db 979 PSVESYEDDESCKSCSYQYPADLAHKLHSANHMDNDGELDTPIYNSLYKSDQELNSGR 1038  
Qy 388 SEQENNEKQMLQSVIPPMFDAQORRVKFINMGLMEDPMKVKYKDRQFMNVTWDEKEI 447  
Db 1039 QSPQNERWARPKHVEIEIKQNEQORARSQNTS-----YPTVISEN-----TD----- 1081  
Qy 448 FKDFIOHPKNGFLIASYLERKSPDCVLYY-----YLTAKNENYKALV 491  
Db 1082 --DKHLKFPQPHFG-----QOECVSPYRSRGTSGSETNRMGSSHAINONVQSLCQ 1129  
Qy 492 RRYTKRRGRNQIARPSQEEKVEKEEDKAEKTEKKEEKKKEEKE----- 540  
Db 1130 EDDYEDDKPTNYS--ERYSEEEQHEEEERPTNYSIKYNEEKHHVDQPIDYSIKYATDISS 1188  
Qy 541 -----KEDSKENTKEKIDGTAETEERQATPRGRKKTANSQGRRRGRTR----- 587  
Db 1189 SQKPSFSKNSAQSTK--PEHLSPSENTAVPPSNAKRQNLRPSSAQORNGTOKGTTC 1247  
Qy 588 -----SMTNEAAAAA----- 602  
Db 1248 KVPINQETIQYCVEDTPICFSRCSLSSADDETCDDTQEADSANTLOTAEVKE 1307  
Qy 603 ---ATEEPPPPPLPPPEPISTEPVETSRWTEEMEYAKKGL--VEHGRNWAALAKMVG 656  
Db 1308 NDVTRSAEDPATEVPAYSONARAKP-----SRLOASGLSESTRHNKAVFEFSSGA 1357  
Qy 657 KSEAOCKNFYNYKRRHNLNLLQHKQKTSRKPREE-----RDVQCESVASTVSAQE 710  
Db 1358 KSPS-----KSGAOTPKSPPEHYVQETPLVFSRCTSV--SSLDSFE 1396  
Qy 711 DEDTEASNEENPEDESEVAVKPS--DSPENA--TSRCNTEPAVELEPTTETAPSTPS 766  
Db 1397 SRSASSVQSPCGSMVSGIISDLDPDPSGOTMPPSRSKTPPP--PPQTVQAKREYPK 1453  
Qy 767 LAVSTPAEDSESVETOVNDISAEABQMDVDOEISAE--GSVC-----DP 813  
Db 1454 SKVPAEAKRESGPKOTAVNAVO--RVQVLPDVDTLLHFATESTPDGFCSSLSALSIDE 1512  
Qy 814 PPATKADSDVVEVRV--PENHASKVEGDKTERDLDRASEKVEPRDEDLVVAQINAQRP 871  
Db 1513 PFIQK-----DVELRIMPVQ-----ENDNGNETESEOPEESNQDKREV-----EKP 1555  
Qy 872 EPOSD--NDSSATCSADEDVD-----GEPEQRMPMDSKPSLLNPTCSILVSSPLK 921  
Db 1556 DSEKDLDDSD-----DDIEILEECIISAMPTKS-----SRKAKKLAQTASKL--PPVPA 1604  
Qy 922 PNPLDLP-----QLQHRAAVIP-----PMVSC--TPCNIPIGTVPVGYAL----- 959  
Db 1605 RKPSQLPVYKLLPAQNRLQAKQHVSTFGDDVPVRYCVETGPIINFSTATSLDLTIESPP 1664  
Qy 960 -----YORHIKAMHESALLEEQRQEQIDLE 986  
Db 1665 NELATDGVRAIGQSEFEKRDRTIPTEGRSTDDAQRGKISSIVTPDLDDNKAEEGDILAE 1724  
Qy 987 CRSSTSPCGTISKSPNREWEVLOPAPHQLITNLPGVRLPTTRTPRPPPLIPSSKTTVAS 1046  
Db 1725 CINSAMPKGSKHKPFYKKTMDVQVQASSTSSGANKNQVDTKKKKTPSPVKPMQNTFYR 1784  
Qy 1047 EKPSFIMGSSIQCTPGTYLTSHNQASVTOETPKPSV-----GSSIS 1087  
Db 1785 TRVRKNTDSKVNVTETFF-----SDNKDSKPSLQTNAKAFNEKLPPNEDVRGTGA 1837  
Qy 1088 LGLPROQESAKSATPLYIKOEESFSPRSQSOPEGLLVAHQEGVYVRGTAQIESGIT-R 1146  
Db 1838 LDSPHHYTPIEGT--PYCFSRNDSLSLDDDDVDLSREKAELKKGESKDEAKVTCTR 1895  
Qy 1147 GTPFSKISVES-----IPSLRGS---ITQGTALPQT-----GIPT--EALVKSGSIS 1188  
Db 1896 PEPNSSOAAASKSOASIKHPANRAQSKPVLOKQPTFFQSSKDGPDGAAATDEKLOLAIE 1955











Qy	168	-DONASPKLSKEELTQSDMRVDRE-----IAKVBEQQILUKKKQQQ-----	208
Db	266	LQRRVSTSR-SQHDILQMCSSRRRGAGSEANLIVAKSWADVYLGAKAQOTQTKVIKHGPQR	324
Qy	209	-LEEEAAKPEPEKPY-----SPPPV-----EQKH-----RSIVQIILYD	241
Db	325	SMKRRQRATPAKPKPVGEVHSQFSTGHANSPTCTIIIGKAHTEKHVHVPARPYRVLLNNEISN	384
Qy	242	ENRKAEEAAHKIPEGILGPKVLEPLYNQPSDTKVYHENIKTNQVMRKLLI-----	290
Db	385	QKMDFKED-----LSGIAEMFKTPVKEQPOLSTCHIASNSENLLGKQFQCTDSEEPPLL	440
Qy	291	-----LFFKRRNHARKQBEQKICORYDOLMEAWBKVKYDIENPNRRKAK-----ESK	337
Db	441	PTSESFGGNVFFSAQN-AAKQPSDK-CSASPLRRQCRQRENGNTAKTPRNTYKMTSLSTK	498
Qy	338	TREYVEKQPEIRKQBEQOERQVQRCAGISATIAEHEISEIIDGLSEQENNE---394	
Db	499	TSD-----TEPESKTYSTVNRSG---RSTEFNIOKLP--VESKEEFTNTEIVE	543
Qy	395	---KOMROLSVIPMMFAEQRRVKFINNGLM---EDPMKVYKDRQFMNVWTDHEKEIF	448
Db	544	CILKRGOKATLL-----QQR-----EGEMKEIERPPETYE---NIELKENDE--	584
Qy	449	KDXFIQHPKNFGL-----IASYLERKSVPCVLYLYLTKKENYKALVRNRYKRRGRNQ	504
Db	585	KMKAMKRSRTGGKCAPMSDLTDLKSLPDTEL-----KDTARGQN	625
Qy	505	IAPSOBEKVEBEEDKAK-----TEKKEEKKDEBEKDEKED-----	543
Db	626	LIQTDQHAKAPSEKGIITKMPQCSLOPEPINTPTHTQQLKASLGKVGKKEALLAVGKF	685
Qy	544	---SKENTKEKDKIDGTAEET---EEREQATPRGRKTANSOGR-----	581
Db	686	TRTSGETHTHRPAGDGAISITFYESPKQIILDPAAVRTGMKKWPRPPEBAQSLDLAG	745
Qy	582	-----KGRITRSMTEAAAAAATAEPPLPPPPPEPISTEPVTSRW-----	628
Db	746	FKELFQTPGPSBEESMTDEKTTKIACKS-----PPPEVDTP-STSKQWPKRSRLR	793
Qy	629	---TEEMEVAKKGLVEHGRNAAIAKAVGTGKSEAQCNFYFNKRRHNLNLLQOHKQK	685
Db	794	KADVEEFLALRKLTPSAGK--AMLTPKPAGDEKDIKAFMGTPVQKLDLAGTLPGSKRQ	851
Qy	686	TSRKPREE---RDVSQCSVASTVSAQDEDEIASNEE---ENPDSVEAVKQPSD	736
Db	852	L-QTPKEKAQALDLAGFKELFQTPG--HTELVAAAGTKTIKPCDPOSDDPVTPTTKQ	908
Qy	737	SPENATSRGNTEPAVELEPTTETAPSTPSLAVPTKPAEDESVEYQVNDISI-----	788
Db	909	RPKRSIRKADVEG--ELLACRNLMPSAGKAMHTPKPSVGEEKDIIIFVGTVPQKLDLTEN	966
Qy	789	-----SATAEQM-----DVDQOEHSAEGSV-----CDPPPATKASVSD	823
Db	967	LTGSKRRRQTPPKEEAQALDLAGFKELFQTPGHTGTEEVAAGKTKMPCCESSPPESADTPT	1026
Qy	824	VEVRPE-----NHASKVEGDGNTKERDLDR--ASEKVEPR	856
Db	1027	STRKQPTPLEKRDVKOKELSALKKLTQTSGETHTYDKVGGEDKSINAFRTAKQKLDLP--	1085
Qy	857	DEDLVAAQIINAQRPQSDNDSSATCSADEVDCEPERQRMFPMWDSKPSLLNPTGSIIV	916
Db	1086	-----AASVTGSKRHPKTKERQAQLP---EDLAGMKELFQTPVCTDKPTTHEKTTKIAC	1135
Qy	917	SSPLKPNPLDLPOLH-----RAAVIPPMVSGTCTPNIP	949
Db	1136	RS--QDPDPVDTPTSSKPSQSKRSRLKVDVEEFPALFKRRTSAGKAMHTPKPAVSGERNIY	1193
Qy	950	--ICTPVSGVALYQ-----RHIKAMHESAL-----LEEQORQOQID	984
Db	1194	AFMGTVPQKLDLTENUTGSKRRLQTPKEKAQALDLAGFKELFQTRHGTESMTNDKTAK	1253
Qy	985	LECRSS-----TSPCGTSSKS	999

Db	1254	VACKSSQPDLDKNPASPSSKRRLKTSIGKVGVBELLAVGKLTQTSGETHTHTPEPTGDGKS	1313
Qy	1000	PNREWEVLQPAHQHIL---TNLPEGVRLPTTRPPPLIPSSKXTTVASEKPSFIMGGS	1056
Db	1314	-----WKAWESPQKQILDASAUSGSKRQLRT-----PKGSEVPEDLAGFI---E	1356
Qy	1057	ISQCTPGTYLTSHNOASYQE--TPKPS-----VGSISLGLPQROESAKSATLPIYIKQ	1107
Db	1357	LFO--TP-----SHTKESMTNEKTTKVSYRASQPDLDVPTTSKQPQKSLRKAD---TE	1406
Qy	1108	EEFSPRSONSQEGLLVRAQHEGVVRGAGALQEGSITRGFTPTSKISVESIFSLRGSIQ	1167
Db	1407	EEFLAFRKQTPSAG---KAMH--TPKPAVGEKDNITELGTVPQK-----1446	
Qy	1168	GTPALPQFG-TPTREALVKGISRMPIEDSSPBKGREEAASKGHVIEGKSGHILSYDNIK	1226
Db	1447	----LDQGNLP-----GSNRRL-----QTKRKAQ-----ALEELT	1474
Qy	1227	NAREGTRSPRTAHEISLRKSYBSEBGNIKQGMBSRESVSPAPLEGLICHALPRGS---PH	1283
Db	1475	GFRELFOPTCDNPTADEKTKKI-----LCKSPQSDPAD-----TPTNTKQRPK	1519
Qy	1284	SDLKERTVLSSIMOGTTPATTESPEDGLKYPQKIKRSPPPIRAPEGALT-----1333	
Db	1520	RSCLKADVEEFL---AFPKLTSPSAGKAMHTPKAAVGEKDNITFVGTPEVKDLGLGNLP	1576
Qy	1334	-----KGKPYDGIITTIKEMGRS---IHEIPRODILTQESRKTPVEVQSTRPIIE	1379
Db	1577	GSKRRPQPKERAKALEDLAGFKELFOPTPGHTEESMTDDKITEVCSKSP-----1625	
Qy	1380	GSISQCTPIKFNNNSGQ----SAIKHNKVSKLITGPSKLSRGMPPLEIVPENIKVVERGY	1435
Db	1626	---QPDVPKPTSSKQRLKISLGKVGKVEEVLPCVKLTQ-----1661	
Qy	1436	EDVKAGETVRSRHTSVSGSPSVLSTLHEAPKAOLSPGCIYDDTSAR--RTPVSVQNTMS	1493
Db	1662	---TSGKTTQT--HRETAGDGKSI--KAFKESAKOMLDPANYGTGMERWPRTPKEEAQSL	1715
Qy	1494	--RGSPPMNRITSDVTIPPNKSNINHERKSVLTPTQRESIPAKSPVP--GYDPVVS--HSPED	1548
Db	1716	DLAGFKELFOPTD-----HTEEST--TDDKTTKACKSPPESSMDTPTSTRRRPT	1764
Qy	1549	PHHRGSTAGEV-----YWSHLPTQLDPAMPFHR-----ALDPAAYLFORQLSP	1593
Db	1765	PLGKRDIYEELSALKQLTQTHTDKVPGEDKGINVFRETAQKLDPAASVTSKSRQ--PR	1823
Qy	1594	TPGYPSQ--YQIYAMENTRQT--ILNDYITSQOMQVNLRPDVARGLSPRQPLGLPY----	1646
Db	1824	TPKGKAQPLEDLAGLKELFQTPCVCTDKPTTHEKTKI---ACR--SPQPDVGTGTIEKP	1878
Qy	1647	PATRGI-----IDLTNMPPTI---LVPHPGGTSTPPMDRITYIPGTOITPPPRY	1693
Db	1879	QSKSLRKADVEESLALURKRTPSVGKAMDTPKPAGGDEKDKAFMGTPVQKLDLP-----	1934
Qy	1694	NSASMPGHPHLLAAAAAER	1753
Db	1935	---GNLPG-----SKRWPTQPKAQALEDL--AGFKELFQTPGTDKTTDEKTTK	1980
Qy	1754	V--RSPSPSVRTQETMLOORPSVFOCTNGTSVITPLDPTAQLRIM--PLPAGGPSISQGL	1809
Db	1981	TACKSPQDPDVTDPASTKQRPK-----RMLRKADVEEEFLALURKRTPSAGKAMDTPK	2032
Qy	1810	PA-----SRYNTAADALALVDAASAAPQMDVSKTTEKSKHEARLENLRS-----	1860
Db	2033	PAVSKDENINTFVETPVQKLDLLGNLP--GSKRQPTPKAEALLEDLVGFKELQTPGH	2090
Qy	1861	SEOOOLEQKTLTEVKRSVQCLYTSSAFPSGKQPQHS--SVVYSEAGDKGPPPKSYEEF-	1918
Db	2091	TEESMTDKITEVSCKS-----POPESFTRSRSKORLKIPLVKVDWKKEFP	2136
Qy	1919	-----LRTRGKTTITTAANFIDVITTRQIASDKARSGQSSDSSSLSSHHR-----	1965



Db 2137 LAVSKLRTSGTQTHTE-----PTGDSKSIKAFKESPKQILDPAASVGTGSRRLRTRK 2191  
Qy 1966 -----YETPSDAIEVIS-----PASSAP-----PQE 1987  
Db 2192 EKARALEDLVDFKELFSAPGHTESMTDKTKIPCKSPPELTDATATSKRCPKTRPRK 2251  
Qy 1988 KLOTYQPVVANKAQNEDPTQYEGPL-----HHYRPOQESPSPOQLPPS 2033  
Db 2252 EVKEELSAVERLTQTSQSTHTRKEPASGDEGIKVLKQRAKKKPNVPEEPSRRPRAPK 2311  
Qy 2034 SOAEGMGVPRTHRLITLADHICOLITODFARNVSSOTPOQPTST----- 2080  
Db 2312 EKAQPLEDLAGTELSETSGTQSLTAGKATKPCESPPLLEVVDVDTASTRHLNTRVQK 2371  
Qy 2081 --FQNSPALVSTPVRTKTSNR--YSPESAQSVHQRGSRVSPENLVKDSGRSPKQSP 2137  
Db 2372 VQKREPSAVFTQTSGETTDADKEPAGEDGKIKALKESAKQTPAPASVI--GSR--RRP 2428  
Qy 2138 ERSVSEPEPIPPQPVVHVKQDSLLLSQSGAEPAEORNDARSQSGISYLPSPFTK 2197  
Db 2429 RAPRESAQAIEDLAGFKDPAAGHTEESM-----TDDTKTKIPCKSS-----PE 2471  
Qy 2198 LENTS-----PMVKSQKOEIIFKLNSG-----GGDSMAAOP-----GTEIFNLPA- 2240  
Db 2472 LEDTATSKRRPRRAQKVEKELLAVGKLQTQTSGETTHDKEPVGEGKGTAKFKQPAK 2531  
Qy 2241 --VTTSGSVSRGSHFADPNASNLGLEDIIRKALMGSDDKVEDHGVVMSQPMGVVPGTAN 2298  
Db 2532 RNDAEDVIGRRQPRAPKKAQPLED-----LASFOE-----LSQTPGTHELAN 2577  
Qy 2299 --TSVTSGETREEDGPPSHSGVCK--PKLISKNSRKSIPGQYGLGTERP 2350  
Db 2578 GAADSFTSAPKQTPDSGKPLKISRRLRAPKVEPVGVVTRDPVKQSKSNTSLP 2633  
  
RESULT 13  
S37431  
N:Alternate names: ankyrin long splice form - human  
N:Contains: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid  
N:Species: Homo sapiens (man)  
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 13-Aug-1999  
C:Accession: S37431; A35643; B39643; A40334; A49462; S14533; S14569  
R:Chan, W.  
submitted to the EMBL Data Library, September 1993  
A:Reference number: S37431  
A:Accession: S37431  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-3924 <CHA>  
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288  
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.  
J. Cell Biol. 114, 241-253, 1991  
A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a  
A:Reference number: A39643; MUID:91302466  
A:Accession: A39643  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2077 <OTL>  
A:Cross-references: GB:X56957  
A:Accession: B39643  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1443,3585-3924 <OTT>  
A:Cross-references: EMBL:X56958  
R:Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,  
Genomics 10, 858-866, 1991  
A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.  
A:Reference number: A40334; MUID:92009921  
A:Accession: A40334  
A:Molecule type: DNA  
A:Residues: 463-474, 'PE', 477-495 <TSE>  
A:Cross-references: GB:N37123; NID:g178647; PIDN:AAA62828.1; PID:g178648

R:Chan, W.; Kordeli, E.; Bennett, V.  
J. Cell Biol. 123, 1463-1473, 1993  
A:Title: 440-KD ankyrin-B: structure of the major developmentally regulated domain and  
A:Reference number: A49462; MUID:94075409  
A:Accession: A49462  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-3924 <RES>  
A:Cross-references: EMBL:Z26634; NID:9406287; PIDN:CAA81387.1; PID:9406288  
C:Genetics:  
A:Gene: GDB:ANK2  
A:Cross-references: GDB:127607; OMIM:106410  
A:Map position: 4q25-4q27  
C:Superfamily: ankyrin; ankyrin repeat homology  
C:Keywords: alternative splicing  
F:2-3924/Product: ankyrin 2, long form #status predicted <MAT>  
F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>  
F:63-95/Domain: ankyrin repeat homology <AN01>  
F:96-128/Domain: ankyrin repeat homology <AN02>  
F:129-161/Domain: ankyrin repeat homology <AN03>  
F:162-190/Domain: ankyrin repeat homology <AN04>  
F:191-223/Domain: ankyrin repeat homology <AN05>  
F:232-264/Domain: ankyrin repeat homology <AN06>  
F:265-297/Domain: ankyrin repeat homology <AN07>  
F:298-330/Domain: ankyrin repeat homology <AN08>  
F:331-363/Domain: ankyrin repeat homology <AN09>  
F:364-396/Domain: ankyrin repeat homology <AN10>  
F:397-429/Domain: ankyrin repeat homology <AN11>  
F:430-462/Domain: ankyrin repeat homology <AN12>  
F:463-495/Domain: ankyrin repeat homology <AN13>  
F:496-528/Domain: ankyrin repeat homology <AN14>  
F:529-561/Domain: ankyrin repeat homology <AN15>  
F:562-594/Domain: ankyrin repeat homology <AN16>  
F:595-627/Domain: ankyrin repeat homology <AN17>  
F:628-660/Domain: ankyrin repeat homology <AN18>  
F:661-693/Domain: ankyrin repeat homology <AN19>  
F:694-726/Domain: ankyrin repeat homology <AN20>  
F:727-759/Domain: ankyrin repeat homology <AN21>  
F:760-792/Domain: ankyrin repeat homology <AN22>  
F:793-825/Domain: ankyrin repeat homology <AN23>  
  
Query Match 3.3%; Score 411.5; DB 2; Length 3924;  
Best Local Similarity 18.7%; Pred. No. 3.9e-08;  
Matches 493; Conservative 357; Mismatches 946; Indels 845; Gaps 114;  
  
Qy 57 LQOQQOQQLRRRPSILSEFHPGSDRQERRTSYEPFHPGSPVDHDSLESKR-----P 109  
Db 1740 LQKRVDEQKGRSKLPKVKGEDVP--KKTHRP--HPAASP-----LSSSERHAPGSPSP 1792  
Qy 110 RLEQVSDSHFORVSAAVLPLVHLPEGLRASADAKKDPAGFGKHEAPS--SPISGQPCGDD 168  
Db 1793 KTERHST-----LSSSAKTERHPPVSPSSKTEKHSVPS--PSAKT 1830  
Qy 169 QNASP-----SKLSKEELIQSDRVDRVRETAQVQOILKKKQOOLEEBAAPPEPEKPV 224  
Db 1831 ERHSPASSSSKTEKHSVSPSTKTERH-----SPVSTKTERHPPVS 1872  
Qy 225 P-----PPVEQKHRSIVQIYDENRRKKAFAHIFBGLGPKVELPLYNPSDTKYVHE 277  
Db 1873 PSGTKDRKPPVSPSGRT-----EKHPPVSPGTERKLPV--SPSGRTDKHQ 1916  
Qy 278 NIKTNQVMRKILIFFRRNRHAR-----KOREKICQRYDQDLMEAWKKVDRIEN- 327  
Db 1917 PVSTAGTKTERHLPVSPSGKTEKQPPVSPTSKTERTETMSYR--ELMKAFSGQDPSKHK 1974  
Qy 328 -----NPRKAKESKREYKQFEIRKQEQO-----ERFQVR-----GQRG 366  
Db 1975 TGLFEHKSQKQKQPKRQKRVKEKGPILTQREAAQTENQTIKRGQLPVTGTAEKRG 2034  
Qy 367 AGLSATATARSE-----HEISEIIDLSEQENNEKOMRQLSVIPPMFADQRRV 415  
Db 2035 VRVSSIGVKKEDAAAGGKEKVLSHKIPVQSVPEESHR-----ESEPVE 2080







Db 3768 VSTPAEEKLYQTP-----TSSRGSGPIIQPEEPSEHRESSPRKTSILVIVESADN 3821  
Qy 2410 Q 2410  
Db 3822 Q 3822  
RESULT 14  
T25752  
hypothetical protein F45E4.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25752  
R:Wilson, R.  
submitted to the EMBL Data Library, September 1996  
A:Description: The sequence of C. elegans cosmid F45E4.  
A:Reference number: Z20082  
A:Accession: T25752  
A:Status: preliminary; translated from GR/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2361 <WIL>  
A:Cross-references: EMBL:U70852; PIDN:AAB09135.1; GSPDB:GN00022; CESP:F45E4.4  
A:Experimental source: strain Bristol N2; clone F45E4  
C:Genetics:  
A:Gene: CESP:F45E4.4  
A:Map position: 4  
A:Introns: 60/2; 111/2; 939/3; 977/2; 998/1; 1021/1; 1125/3; 1140/3; 1157/2; 1173/2; 131/2  
Query Match 3.2%; Score 409.5; DB 2; Length 2361;  
Best Local Similarity 18.6%; Pred. No. 2.5e-08;  
Matches 523; Conservative 374; Mismatches 975; Indels 943; Gaps 138;  
Qy 79 SDRPQERTSYEPHPGSPVDHDSLESKRPRLEQVSD--SHFORVSAAVLPVHPVPEG 136  
Db 22 SDRPSALAME-----TTPSPASPSL--LRKESEDRGDIALLRRSGA-----D 65  
Qy 137 LRASADAK-----KDPAFGGKHEAPSSPISGQPCGDDQNASP-----SKLSKEELIQSM 185  
Db 66 SRASNDSSASRLPTALVGLSEAEKQHIMSVMSNRNTSPWTSRRCSALOMLPVDNL 125  
Qy 186 DRVDRE-----IAKVEQO-----ILKLLKKQ-----QLEEEAAKPPPEKP 222  
Db 126 SEAEKEHQTILERAESKTPFMKVPKMKQISSRTESTNSRVSEGDIEEVEVQRRKT 185  
Qy 223 VSPPPVEQKHSIVQIYIDENRKAEEAHKIFELGPKVEL----PLYNQPSDTKVYHEN 278  
Db 186 IEEPIVEIPSAVT-----PRNLRV-----IPPIAISHTPPHSAKTDGSRHSS 232  
Qy 279 -----IKTNVMRKKLILFKRRNHARKQREOKICORYDOLMEAWKKVDRIENNP 329  
Db 233 GSSAHSQFGSTPSISGFK--IPFDK---AKTATETLVKEIKDEVIVEYDK--DKTEKP 285  
Qy 330 RRKAKESKTREYKQPEIRKQREOQORFORVQORGAGLSATI-----ARSEHISEIID 385  
Db 286 EPNVSNELTAE-----EFEHRRINEM-----AGIDEIQQPPPIAQRERKSSVVS 330  
Qy 386 GL-----SEQENNEKQ--NRQLSVTPPMFADAEQRVKFINNGLMEDPMKVYKD 433  
Db 331 GLKNMFGVGHDESELTTEEKEHTRMS-----LMAEKM----- 364  
Qy 434 RQFMNVWTDHEKEIFKDFKFIQHPKN--FGLIASYLERKSVPCVLYLYLTKKNENYKALVR 492  
Db 365 -----DEELDEQ--SKPKSTFGL-----KNF----- 384  
Qy 493 RNYGK-----RRGRNQIARPSQEEKVEEKEEDKAETKEKKEEED 539  
Db 385 --FCKATQSVMHARDSVYKVNQ--SKQSLSGLTQELDNIAQATSAQOESKQELTQD 440  
Qy 540 EKEDSKENTKEKDKIDGTAETEETEREQATPRGRKANSQGRKRITRSMTNEAAAAA 599  
Db 441 EIDHIA-----RINAAAEDEFQPAVN-----MNOG-----LSQAEKDHIARIE 479

Qy 600 AAAATEEPPLPPPPPEPISTEPVETSRWTEEMEVAKKGLVEHGRNWAALAKVMGTKSE 659  
Db 480 AWAADDQSOKIIVPPPSIIQKD-IELSM--EEMD-----HIARAAAD----- 520  
Qy 660 AOCKNFYNYKRRHNLNLLQHQKQTSRKPREEDVSQCE-----SVASTVSAQED 711  
Db 521 ---EDF-----THPVKGAVALIYDENVSRDGDCAIDRFPASTATPVFAQPS 562  
Qy 712 EDIEASNEENPEDSEVAVKPSDESPENATSRGNTPEA--VELEPTE-----TAPS 762  
Db 563 E-IELS-EEEREHISRIAAAEEDFN-----TPYVSTHTQOIELELTEEKEHIAIRIAM 616  
Qy 763 TSPSLAVPSTKPADESVETQ---VNDISIAETAEQMDVDQOEHSAEAGSCVCDPPPAKKA 819  
Db 617 ASEDINAPSPFIQOQRVATPVPIYHVEPTLSOEI--HIARITAM-----ATE- 667  
Qy 820 DSDVVEVRVPHENHASKVEGDNTERDLDRASEKVEPRDEDLVAAQIINAQPEQSDNDS 879  
Db 668 DNTDLQT--LPTPOVKQNEPELSQE--EIDHITRIAAAMANEDEFGMYPYVSEHPAPV--ES 722  
Qy 880 SATCSADEDDVDGEPRQRMFPMDSKPSLLNPTSLVSSPLKPNPLDLPO--LOHRAAVI 937  
Db 723 ELT---QEELDHIAKITGMASNDIS--TLPPPTG-----KPSETSLTQEELDHIAKIA 770  
Qy 938 PPMVSTPCNIPICITPVVSGYALYQRIKAMHESALLEBQRQEQIDLECRSSTSPCGTS 997  
Db 771 E--MASAEYDVP-----KIFPELPELTQEELDHIAKITAMAA--- 805  
Qy 998 KSPNREVEVLQAPHLITNLPVGVRLTTPRTPPPPLPSSKTTVASEKPSFIMGSI 1057  
Db 806 -----QDVQLPATQRSITH-----KVSLLTPP--PPP----- 830  
Qy 1058 SQGTPGIYLTSHNQASTQETPKPSVGSISLGLPQQQSAKSATLPY-----IKOEE 1109  
Db 831 -----SKHFQDLTQEE-----LHIAERIAEMADMDYNTPTAEPVQDEE 870  
Qy 1110 FSP--RSONSOPEGLLVRAQHEGVVVRGTAGAIQEGSIRGTPTSKISVESIPSLRGSTQ 1167  
Db 871 EEPITESGDSATGADIFDEQEDASSGASGDFDNNAQ--VLTSGFSPDRVTSAPALDT 928  
Qy 1168 GTPALPQTGITEALVKGSIISRMPIEDSSPEKGREEAASKGHVYIEGKSHILSYDNTKN 1227  
Db 929 EEOGP-----IMAKTVSTPTPSADSMASRKSSSE-----YD-----IRSIEIR 968  
Qy 1228 AREGTRSPRTAHEIS-LKRSYVESVGNKQMSRESVPSAPLEGL-----ICRALPGRS 1281  
Db 969 EESDIGNKWEQELSFMRQSIHDEEDV--GHEINTDVEEPFLEYEDQLHLEGIDVES 1026  
Qy 1282 PHSDLKERTVLSG---SIMQG-----TPRATTESFEGLKYP 1315  
Db 1027 QHHEFTTSSAFFGTGRSIGEGEKRYGDDAVEQOKLENYEEEEKTKSSREAFDDGFETQ 1086  
Qy 1316 KQ--IKRESPPIRAFEAGAITKGP-----YDGIITIKEMGRSH-EIPRODILFQE 1363  
Db 1087 REESLRAQOTPIDSLGSRMLKRPNGFLSNIANDAINKAKEAGSOIQAAVPIKPSSSSS 1146  
Qy 1364 SRKTPVVVQSTRPIEGSISOGT--PIK-----FD 1391  
Db 1147 NIVNNVFFSSK-----SSTSLGTSAPTSPQIGIPMDGLSEERKQIMSVMAAADFD 1202  
Qy 1392 NNSGSAIKHNKVLITGPKSLSRGMPPLEIPVENIKVVERGKEDVKA-----GETVR 1445  
Db 1203 DS-----VNNVKESTGSSNIPAGM-----EDLSEAREKIMSVMANEMENGARFP 1249  
Qy 1446 SRHTSVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSARTPV-----SYQNTMSRG 1495  
Db 1250 PPSQIPIRSPSVMSSSI---MSELPPGLDLDSDDERMKIMAYMAEADMQNVKPIARG 1305  
Qy 1496 SPMNRTSDVTIPN-----KSTNHERKSTLTPTQRE 1527  
Db 1306 PPPMPPSTSM--IPPGMEGLSEERQKIMSVMANAETDSSSVITSRQPSRSPSVARMQPK 1364



Qy 1528 SIPAKSPVPGVDP-----VSHSPDPHHRGSTAGEVYWSHLPTQLDPAM 1572  
Db 1365 LMPQQAIPPIPPGLEGLSDEERHTIMSVMAEAEFEESR--SQVPSRQPSRSPSFVNQQ 1422  
Qy 1573 PFHRALDPAAYLFFQRLSTPGY-----PSQVQLYAMENTROTILNDYITSOQMVN 1627  
Db 1423 SFH-----PIPSFEPIVPPGLEDL--SEERQKIMS-----VMNA 1456  
Qy 1628 RPDVARGLSPREQPLGLPYPATRIIGLIDLTNNPPTILVPHPGGTSTPPMDR---ITYIPGT 1684  
Db 1457 EVEESRQLPSRQPSRSPVA-----MIQAPVPIIP-SCMEDLPEAEROKIMSVMAEA 1509  
Qy 1685 QIT--FPPRPYNASMSPGH--PTHAAASAERER--ERERERERERERIAAASDLYL 1738  
Db 1510 EIQNFVPKSRSPSNYSQVPPVIPHGLEDLSEARQKILSVMAEAEIDSAKIPSRSTSYS 1569  
Qy 1739 R-----PGSEQP-----GRPGSHGYVRSPSPSVRTQETMLQQRSPVFGT 1778  
Db 1570 MPPPLQMSQPEITITGLEHSEADMEFGDRSSRSHOVIPPGLENLSEERQOIMSVMAHA 1629  
Qy 1779 NGTSVITPLDPTAOL-RIMPLPAGGSPISQGLPASRYNTAADALAALVDAASAPQMDVS 1837  
Db 1630 ELESII-----PSGHIDQISLPRGHTGK---PAGIVN--EDEL-----FETERKQREES 1675  
Qy 1838 KTKESKH-----EAARLEENLRSR-----1856  
Db 1676 PTRESGYATSYREELAMGDEERMDGLEDIIRIREGARSRRSDRDLHRRREEDPEVH 1735  
Qy 1857 -----SAVSEQQLEKQTLVEKRSV--OCLYTSSAFP-----1888  
Db 1736 TPEESSTAVTDPVSPVTENVPEKQDFDFTYSDSRFAEIVQMDEEAGSLQOKV 1795  
Qy 1889 SGKPO-----PHSSVYVSEAGKDKGPP--PKSKYE-----EELRTGK 1924  
Db 1796 DEKPRMWTVDGDESELPHQDFVNEPTTKTSDFPKDETFEVPKPSIQIRIVTKN 1855  
Qy 1925 TTITAAINFIDVITRIQIASDKDARERGSSQSSSSLSH-----RYE 1967  
Db 1856 HDVDMDEIDYDVIATEAPSSVSQRR--QPVDSETSVKSRTIQRGTPKPPMIKITVEE 1912  
Qy 1968 TPSDAIE-----VISPASSAPPQEKLOTQVPEVVRKANAENDPTQYEGPL 2014  
Db 1913 TKSDSDEESCEDDEEYDPRVVAAPTAPTFE-----EVENERIKQ-----1954  
Qy 2015 HHYRQPSRSPQOOLPSSQAEGMGQVPRTHRLITLADHICQIITQDFARNQVSSQTPQ 2074  
Db 1955 -----EELGKVLQI-----NAFG-----EVANDEFQVQAKTTTSQ 1987  
Qy 2075 QPPTSTFQNSALYSTVPTVTKTNSRYSPEQAQSVHHQRGSRVSPBNLVDKS-----RGS 2131  
Db 1988 TPSTSTKPTVTPAKRSDPI-----PIAPSQRKSEIEE-----RIRTEALEEEBFYRHG 2037  
Qy 2132 RPK-KSPERSHVSSEPEYPISPQVPPVHVKQDSLILLSQGAEP-----EQNRDARS 2184  
Db 2038 NPFLESPEDEVSIN-MEDVYEAIRLYESANO-----TMRPGPVYITIDESEDDGTL 2092  
Qy 2185 PGSIYLPSPFTKL--ENTSPWAKSKKEIFR-KLNSGGGSDSMAAQAQGTIEFNLPV 2241  
Db 2093 SNSESRLVAREKRLMDKTAUDSLMAKYQKKKKVQAKQTITASSTSVTATPAYAIINSD 2152  
Qy 2242 TSSGVSRRGSHFADPASNGLGDI-----IRKALMGSPDKVEDHGVMVMSQ-----2288  
Db 2153 KTS-TRTTDSNSYFETTKNIPALEIKDPKKDIPPEISASIDKTMAEVDALLGQVYTNKA 2211  
Qy 2289 -----PMGVPGTANTSVTSGETREEDGDPSPHSGGVCKPKLISKNSRKS 2337  
Db 2212 IPNLICFQOSNFNHPPTSSASTATDDL-----ILLKNNSSSPSF 2253  
Qy 2338 --PIPGQGLYTERPSSVSSVSHSGDYH-----RQTPGWAWEDRPSSTGSGTFFY 2385  
Db 2254 LLPLQ-SSVLG-----SOLDSVRDDNERNENETSPRGLKRSFGML--LPSPSTSTSI 2304  
Qy 2386 NPLWMLSSPTPTPIACAPSAVNOAAPHQONRIWEREAPL-----LSAQYETLSD 2437

Db 2305 -----MPPTAASVGAAGATTATSMFGGISVADPPPTMDGLTNSYKWLQN 2349  
RESULT 15  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N;Alternate names: alpha-NAC protein  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
C;Accession: T30826  
R;Yotov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A;Title: Differential splicing-in of a proline-rich exon converts alphanAC into a mus  
A;Reference number: Z20889; MUID:96312450  
A;Accession: T30826  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2187 <YOT>  
A;Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1  
C;Genetics:  
A;Gene: Naca  
A;Map position: 10  
A;Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
A;Note: differential splicing converts alphanAC into a tissue-specific DNA-binding ac  
C;Keywords: alternative splicing; DNA binding; transcription factor  
Query Match 3.2%; Score 408; DB 2; Length 2187;  
Best Local Similarity 20.5%; Pred. No. 2.6e-08;  
Matches 404; Conservative 261; Mismatches 780; Indels 524; Gaps 92;  
Qy 504 QIARPSOEKVEEKEDEKAEKTEKKEEKEDEKDEKDEKENT-----KEKDKIDGTA 558  
Db 497 QAGLPTKRTDQLPLALIKESPPSSQASASLEVLSEDTVTYKKTGGPVPVRAIAGVA 556  
Qy 559 EETEEREQATPRGRKTAN--SQRRKGRITRSMTNEAAAAAATAAEEPPPLPPPPPE 616  
Db 557 TTTSLRADSPPAVIRADSCVSPNTVSQPLKRSVTDPAAPRTAKNTA-----PSTTSPLV 611  
Qy 617 PISTE--PVETSRTEEMEVAKKGLVEHGHENWAAIAKMGVTKSEACKNFYF-----667  
Db 612 PLASEGCPVASS-----MALSPQN-----ASVETALALSPETPKSVPPFPDPLAE 657  
Qy 668 -NYKRRINLNLQHQKTSRKPREERDVSQ-----C-----ESVASTVSAQEDIEDASN 718  
Db 658 ISFSNARKVDVAVSHWESSGSSRGHPDASVTAKGTVVCLADSLDTSVSASKGALSAS 717  
Qy 719 EENPEDE-----VEAVKPSD--SPENATSRGNTPEPAVELEPTETAP-----ST 763  
Db 718 SPLYPLEVSEFLPEAGLAVQGPGLSLNKLSPTPPSKSGAPVPSTGAPSPKGPAPVPTESS 777  
Qy 764 SPSLAVPS-TPAEDESVEITQVNDISIAETAEOMD--VDQOEHSAEESGVCDPP-----814  
Db 778 ISSQVFAELIPSPQKTEVTASRLISAVOSPKVDPTMSOVPTPSPKRTSATAPKDTSA 837  
Qy 815 -----PATKADSDVDEVVRVPHENHASKVEGDNTKERDLORASEKVEPRDLDLYVAQOI- 866  
Db 838 TSLSKSVPAVYTSLSPPKAPVAPSNEATVPTPE--IPTSLKNALAAATPKE---TATSIP 892  
Qy 867 NQAPPEPOSNDSSATCSADSDVDGEPERQRMFMDSKPSLLNPTGSLVSSPKLPN--P 924  
Db 893 KVTSPSPQKTPKSVS-----LKGAP-----AMTSKKAT-----ETAASKDYSPSQFP 934  
Qy 925 LDLQLQOHRAAVIPPMVSCYPCNIPIGTPVSGYALYORHIKAMHESALLEERQORQDID 984  
Db 935 KEVPLLOH-----VPP-----TSPPKSPVSDTLSG-ALTSPPPKG---PPATLAB-----975  
Qy 985 LECRSSTSPCGTSKSPRNEWELQAPAPHLITNLPGLTTPTRTPRPPPLIPSSKTTV 1044  
Db 976 -----TPIYPKSPKPAASKKTPA-----TPSPEGV-----TAVPLEIPCSKKAP 1016  
Qy 1045 ASEKPSFMGSGISQGTGTYLTSHNQASVTOETPKPSVSGSISL--GLPROQESAKSATL 1102



Db 1017 KTAAPKESATSSSRAPKT-----AVSKEIPSKGTAVPLEISLP-LKETSAT- 1066  
QY 1103 PYIKOEESPRSONSQPELLVRAQHEGVGRGTAGAIQ--EGSITRGPTSKISVESIP- 1159  
Db 1067 -----PGEKASAPKSKPK-----TAGPKETPPGGVAVPPEISLPPKETPQ 1108  
QY 1160 -----SURGSITOGTPALPOTGIPTALVKGISRMPIE-DSSPEKREAAKSGHVIY 1212  
Db 1109 NATPNESLAASQKRS-----PKTSVPKET-PPGGVTAMPLEIPSAQKAPKATVPKQITP 1164  
QY 1213 EGKSHIILSYDNINKNARCTGTPRTAHEISLKRVSVEGNIKQGMRSRESVPVAPLEGL 1272  
Db 1165 E-----DAVTILAGSLSPKASKTA-----APKEAP-ATPSPGV 1198  
QY 1273 ICRALPRGSHDLKERVLSQIMOGTPRATTESFEDGLKVPKQIKRESPIRAFECAI 1332  
Db 1199 I-----AVSGEI-SPSKTKST-----AAPKENSATLPPKRSPTAA 1235  
QY 1333 TKGPK-----YDGITTIKMGRSIHEIPRODILTQESRKTPEVVQSTRPIEGSISQGTPIK 1389  
Db 1236 PKETATPSSEGTAV-----PSEISPSPT-----PASKGVPT 1269  
QY 1390 FDNNSGQSAIKHNKSLITGSKLSRGMPPLEI-----VPENIKVVERKYEDV-----KA 1440  
Db 1270 LTPKGAPNALAESPAS-----PKKVPKTAAPETSTTPSPQKIPKVPAGKEASATPPSKKT 1325  
QY 1441 GETVRSRHTSVSSG-----PSVLRSTLHEAPKAQLSPGIIYDDTSARRTPVSYQNTMSRGS 1496  
Db 1326 PKTAVPKTSAPSEGTAVPLEIPSPKAPKT-----AAPKETPAP-----S 1368  
QY 1497 PMNNTSDVTIPP-----KSTNHERKSTLTPTQRESIPAKSPVPGVDVPSVSHSPDPH 1550  
Db 1369 PEGATTAPVQIAPPSPRKSCKKAGSKETPTTSPGCVTAAPLEIPISSKKTSMASPRE-- 1426  
QY 1551 HRGSTAGEVYVSHLPTQLDPAMPFHRALDPAAYLFORQLSPGTPGYPYSQYLYAMENR 1610  
Db 1427 -----TLVTPS-----SKLSQTVG-----PK 1443  
QY 1611 QTIINDYITSOOMVNLRPD--VARGLSPREQPLGLPYPATRGIIIDLNTNPPITLPHP 1667  
Db 1444 ETSLEG-----ATAVLEIPSPSHKAPKTDVDPKQVPL-TPSPKD-----APTLEASP 1490  
QY 1668 GG-----TSTPPMDRITYIPGTQITPPRPYNASMSPGHPTHLAAASAERER----- 1716  
Db 1491 SSPKAPKTAAPPSEVTVTPPERPATPKAGSTASKVPVPAETOEVAVSSRETPTVPA 1550  
QY 1717 -----EREKERERERIAAASDLYLRG-SEQCPRGSHGVRS-----PSPSVRTQ 1764  
Db 1551 VPPVKNPSSHKTSTIKLEKAPATLPPSPKSPKIPSKKAPRTSAPKEPPASPIKPV 1610  
QY 1765 ETML-QORPSVFOGTNGTSVITPLDPTAQLRIMPLPAGGSPISOGLPASRYNTAADALA- 1822  
Db 1611 TSLAQTAPPSLQAPSTTI-----PRENL-----AAPAV-----LPVSSKSPAARAS 1655  
QY 1823 ALVDAASAPOQ--DVSKTESKHEAARLEENLRSAVSEQQOLEOKTLEVEKRSYQC 1880  
Db 1656 ASLSPATAAPQATPAKEATTIPSCKKAATETPIETSTA-----PSLEGAPKETSETSVSK 1710  
QY 1881 LYTSAPSPGAPQPHSSVYSEAGDKGPPKSYEEELRTRGKTTTAAANFIDVITRO 1940  
Db 1711 VLMS-----PPKASSSKRASTLPLATLPLSLKEASVLSPSTA 1747  
QY 1941 IASDKARERSQSSDSSLSHRYETPDSAEIVSPASSPAPQEKLOIYQEVVKAN 2000  
Db 1748 TSSGKDSHI--SPVSDACSTG-----TTPQASEKLPSSKKGPTAFTTEMLAAPAPESALAI 1800  
QY 2001 QANDPTROYEGHLHVRPOQSPSPQOOLPPSSQAQGMQGVPRTHRLITLADHICOLIT 2060  
Db 1801 TA-----PIQKSGANSNSASSPKCFDPSKDKTKGLPNAVALAPQTPVPEKDTSKAIEILL 1857  
QY 2061 QDFARNOVSGTPOOPTSTFTQNSPSPALVSTPVTKTSNRYSPESQAQSVHHQRPGRSVS 2120

Db 1858 VSPAKGSDCLHSPKGPVSGQ-----VATPLAAFTSDKVPPEAVSASV----- 1899  
QY 2121 PENLVDKSRGSRPGKSPERS-HVSSEPYEPISPPQVPPVHKEKQDSLILLISQRGAPAEQR 2179  
Db 1900 -----APKAPAASLTAPSPVAPL-PPKOPLESAPGSVLESPSKLPVPABED 1947  
QY 2180 N-----DARSPGS-----ISYLPSPFTTKLENTSPWYKSKQEIFRKLNSGGG--DS 2224  
Db 1948 ELPLPILPEAVSGGEPFQPIILNMPA--PKPAGTAPAPASAKQPVLK--NNKSGSTESDS 2003  
QY 2225 DMA-----AAQPGTEIFNLPAVT-----TSGSVSSRGHSIFADPA-SNLGLEDI--- 2266  
Db 2004 DESVPELEQDSTOTATQOAOALAAAEIDEPVSKAKOSREKKARKAMSKLGLRQVTGV 2063  
QY 2267 ---IRKALMSFDDKVEDHGVMSQPMGVPGTANTSVTSGETRREE 2311  
Db 2064 TRVTIRKSNILF-----VITKP-DVYKSPASDTYIVFGEAKIED 2102

RESULT 16  
T20532  
hypothetical protein F07A11.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20532; T27777  
R:Palmer, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: 219287  
A:Accession: T20532  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <WIL>  
A:Cross-references: EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone F07A11  
R:Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: 220417  
A:Accession: T27777  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <WI2>  
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone 2K20  
C:Genetics:  
A:Gene: CESP:F07A11.6b  
A:Map position: 2  
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3

Query Match 3.28; Score 405; DB 2; Length 2722;  
Best Local Similarity 18.18; Pred. No. 4.4e-08;  
Matches 442; Conservative 340; Mismatches 823; Indels 836; Gaps 111;

QY 3 SSGYPNQGAFSTEQSRYPHPSVQYTPNTRHQEFAVPDYRSSHLEVSQASQLLQ0000 62  
Db 687 TSGRP-----SLDESR-----TNRLSFDSTHHAELA---QRSHSLCIGPMTPTPTPTS 733  
QY 63 QQLRRRSLLESEFH-PGSDRPQER-----RTSYEPHPGSPVD-HDSLESK-RPRLEQ 113  
Db 734 Q-----PLLNVNTHLPQTSGTSGGITTTPKSSQPP--PLMSPVSRHNSMSSTGRP----- 782  
QY 114 VSDSHFORVSAVLPLVHPLPEGLRASADAKKDPAGFGKHEAPSPISGQPCGDDONA-- 171  
Db 783 -----ASIQTLRH-----QSVFFPDVSIPIPPPI--PTHEMMAPR 817  
QY 172 --SPSKLSKEELIQMSRDVREIAKVEQIILKKKQQOLEEAAKPEPEKPVSPPPVE 229  
Db 818 GTPPSRSSETMVLRL--SPFCTPIQNLITM-----PIVPPP-- 853  
QY 230 QKHSRIVQIYDENRKAEEAHKIFEGLPKVEL---PLYNQSDTKVYHENITKNQMR 286  
Db 854 ----HLIAATSTGTHSVSSSAHST-----PRHSISGTPVHCEPS-----NSKTSQ--- 894











```
Db 2016 DSAVDLSLKHYSILGFADGRYLQGLQYSGFTDLRHPTDLLSLPLRLRYSSVSNYISD 2075
QY 1505 VTIPPNKSTNHERKSTL-----TPTQRESIPAKSPVPGVDPVWVSHSPDPHHRGSTAGEVY 1560
Db 2076 HRYGPRGDAVGQFASLAQYSAATTAREISRCALNSM-----QYGGRHGSGGGPDL 2129
QY 1561 WSHLPTQLDPAMPFRALDPAALAAAYL---FORQLSPTPGYPSPQYQLYAMENR-OTILND 1616
Db 2130 VOYQP-QHGPCLSAQOGLAPLRLSGLGNPTYPEGSPGNLAQYGAASQATAVRQLLPS 2188
QY 1617 YITSO---QMOVNLRDPVAVGLSPRQOPLGLPYATR-----IDLTN 1657
Db 2189 TATVRAADGMIYSTINTPIAATLPTTPQASVLRPMVRGMYRVPVSGVTAVALPLTSLR 2248
QY 1658 MP-----PTILVHPG-----GISTPPMDRITITPGTQIT-----FPPRPYNAS 1697
Db 2249 VPMIAPRVPLGAGLYRYAPRPFIASSVPPAEGFVYLKGRPAATKASGAGGPPRPPLPAG 2308
QY 1698 MS-----PGH-PTHAAAAASA-----1712
Db 2309 VAREEPSTTAPAVKEAPVAPAPAPAPPPGQKPAGEAAGSGSVLSRPASEKEEAS 2368
QY 1713 -----ERER-----ERER-----EKERERETIAAASDLYLR 1739
Db 2369 QEDRQRKQEQQLLERERERERERERERERERERERERERERERERERERERERERERER 2428
QY 1740 P-----GSEQPRGSHGYRSPSPSVRTQETMLQPPSVFGTNGTSVITPLDPTAQL 1793
Db 2429 TIKOHVLOQOEEERQAFALQREOLAQOQLQOQLOQO-----LQLEEQKQR 2480
QY 1794 RIMPLPAGGPGISOGLPASRYNTAADAL-----AALVDAAS-----APOMDYSKTKE 1842
Db 2481 QKAPFPATCEAPSGPPPAATELAQNGYWPPLTHAFIAVAGTEGPGQPREPV----- 2534
QY 1843 KHEAARLEENLRSAVSEQ---OOLEOKTLEVRKSRVQCILYTSAPFSPGKPPHPS----- 1896
Db 2535 -----LHRGLPSSASDMSLQTEQWEAGRSIGIKRHSMPRLDACEPSGPDSTVRR 2588
QY 1897 --SVVYSEAGDKGPPPKSRVEELRGTGTTTAAFNIDVIITROIASDKD-----A 1947
Db 2589 ADSVQTDDDEEG-----RY-----LVTRRRTRRSA-----DCSVQTDDEDNADWQPVR 2637
QY 1948 RERGSSQSDSS-----SLSSHRYETPSDA-----IEVISA- 1979
Db 2638 RSLRSHSDSGSKHDATASSTTAATARAMSVGIQTITSDSVQTEPQLRVSPAI 2697
QY 1980 --SSPAPPQEKLOTYQPEVKVANAEN-----DPTROYEG-----PLHHYRP--- 2019
Db 2698 HITATDPKVEIVRYISAPEKTGRGESLACQTEPDGAQGVAGPQLIGPTAISPLPGIQ 2757
QY 2020 -----QOESFSP-----QOOLPPSSQAEGWG-QVPRTHRLI-----TLADHICQI 2058
Db 2758 IVPGALGRFEKKKPDPLEIGYQAHLPESLSQLVSRQPPKSPQVLYSPVSLPSPH--RL 2815
QY 2059 ITQDFARNO-----VSSQPOOB-----PSTFONSPSALVSTPVRTKTSNRY 2101
Db 2816 LDTSFASERLANKAHVSPQKOFIADSTLRQOTLPRPKMTLQRS-----LSDPKPLSPTAEE 2871
QY 2102 SPESQASVHHQR-GRKVS---PENLVKSRGRSPKSPERSH-----VSSPEPYEP-- 2149
Db 2872 SAKERFSLYHQOGLGSGVSLPPLNGVLVRKVRTLPSPPEEAHLPLAGQVPSQLYASL 2931
QY 2150 -----ISPPQVPPVH-----EKQDSLILLSQGAEPAPQRNDARSFGSISY 2190
Db 2932 LQRLGAGTTVPATKASULLRELDRLKLVEHSTKLRKKQALDEEKEIDAK-----LKY 2987
QY 2191 LPSEFTKLENTSPMVSKKQIFRKLNSGG-----GDSOW-----AAQOPGTEIFNLP 2239
Db 2988 LELGITORKES--LAKDRGGRDYPLPGLGCHRDYLDSDSELNQLRLOGCTTPAGQYVDYP 3045
QY 2240 AVTTSGSVSRGHGFADPASNIGLEDIIRKALMGSDFKDVEDHGVMSQPMGVVPGTANT 2299
```

```
Db 3046 A-----SAAVYPATPSG-----PTAFQOPR-FPPAAPQY 3072
QY 2300 SVVTSGTTRREGDPSPHSGGVCKPKLISKSNRSKSPGOGYLGTERPSSVSSVHSE 2359
Db 3073 TAGSSGPT--QNGFPAHQ-----PTYTGSTYPAPYP-PGTGYPAEGLPSPAPAHPT 3124
QY 2360 GDYHROTTPGAWEDRPSSTGSTOPFY-----NPLTMRMLSTPPT--PIACAPS 2406
Db 3125 GHYAATP-----MFTTOSAPFPVQADSRAAHOKPQTSIADLEKQVPTNYEVIGSPA 3177
QY 2407 AVNOQAAPHQNRWIEREPAPLLSAQYE 2433
Db 3178 VTWSSAPPETG-----YSGPAVSGSYE 3199

RESULT 18
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
C:Cross-references: EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 3.2%; Score 404; DB 2; Length 3256;
Best Local Similarity 18.1%; Pred. No. 6.1e-08;
Matches 516; Conservative 368; Mismatches 1083; Indels 882; Gaps 128;

QY 46 SHLEVSQASQLLQOQOQOQLRR-----RPSLLSEFHPGSDRPOERTSYEPFHPGP 96
Db 482 SVDINNFGSDSINESGIPLKRVRVSFGGHLRPELFDENLPPN-----TPLKRG 531
QY 97 SPVDHDSL-----ESKRPLEQVSDSHFQ-RVSAAVLPLVHPLPEGLRASA- 141
Db 532 APTKRSLVHMTTPVLKIIKEQPPSGKQESGEIHVEVKAQSLVTSPPAPSRKTPVA 591
QY 142 -----DAKKDPAFGGKHEAPSSPISQPCGD--DQNASPSKLSKEELIQSDMRVDRE-- 191
Db 592 SDQRRRSCKTAPASSSKSQSTVEPKRGGERVATCLOKRVLSR-SQHDILQMICSKRSGA 650
QY 192 -----IAKVEQOILUKKKQOQ-----LEEAAKPPPEKPV----- 223
Db 651 SEANLIVAKSWADVVKLGAKQTQTKVIKHPQSRMNRQRPRATPKPKPVGEVHSQFSTGH 710
QY 224 --SPPPV-----EQKH-----RSIVQIYIDENRKKAEAAHKIFEGLPKVELPLYNQ 268
Db 711 ANSPCTIIIGKATHEKRVHVVPARYVRLNNFISNOKMDFED-----LSGIAMFKTPVKEQ 766
QY 269 PSOTKVVHENIKTNQVNRKKLI-----LFFKRRNHARKQREOKIC 308
Db 767 POLTSTCHIAISNENLKGKOFQDTSGEPLLTSPESFGNVFFSAQN-AAQOPSK-C 824
QY 309 QRYDQLMEAWKKVDRIENNRKAK-----ESKTREYKEOFFEIKRQEQQERFORVGQ 364
Db 825 SASPLRRCQIRENGVNAKTPRNTYKMTSLETKTS-----TETEPSKTSTVNR 874
QY 365 RGAGLSATIAARSEHETSEITDGLSEQENNE-----KQMRQLSVIPPMDFDAQRRVKFI 418
Db 875 SG---RSTERNLOKLP--VESKSEETNTEIVECILKRGOKATILL-----QQR- 918
QY 419 NMGLM---EDPMKVYKDRQFMNVWTDHEKEIFKDFKFIQHPKNFGL-----IASYLERKSV 471
```



Db 919 --EGMKELERPFETYKE--NIELKENDE--KMKAMKRSRTWQCKCAPMSDLTDLAKSL 970  
Qy 472 PDCVLYLYLTNNENYKALVRNGKRGRRNOQIARPSQEKVEEKEDKAEK-----524  
Db 971 PDELM-----KDTARGONLQOTDHAHAKAPKSEKGIKTKMPCOSIQ 1011  
Qy 525 -----TEKKEEKKDEEKED-----SKENTREKDKIDGTAET---E 562  
Db 1012 PEPINTPHTKQOLKASLGKVGKRELLAVGFTYTSGETHTHREPAGDKSIRTFKES 1071  
Qy 563 EREQATPRGRKANSQGR-----KGRITRSMNEAAAAA 601  
Db 1072 PKQILDPAARTVMKKMPTKPEEAQSLDLAGKELFQTPGSESMDEKTTKIACKS 1131  
Qy 602 AATEPPPPPLPPPEPISTEPVETSRW-----TEEMEVAKKGLVEHGRNWAATAK 652  
Db 1132 -----PPPEVDTPSTKQWPKRSLRKADVEEFALRKLTPSAGK--AMITP 1177  
Qy 653 MVGTSEACQKNFYFNKRRHNLONLQOHKQKTSRKPREE-----RDVSQCESVASTVSA 708  
Db 1178 KPAGGDEKIDAFMGTPVOKLDLACTLPGSKROL-QTPKEKAQALDLAGKELFQTPG- 1235  
Qy 709 QEDEDIASNEE-----ENPDESEVAVKPSEDSPENATSRGNETPAVELEPTTETAPST 763  
Db 1236 -HTEELVAAGTKTIPCDSPQSDPVDPTSTKQRPKRSIRKADVEG--ELLACRNLMPSA 1292  
Qy 764 SPSLAVSTKPAEDESVEQVNDSI-----SAETAEQM-----DV 798  
Db 1293 GKAMHTPKSPGEEKDIIIFVGTVPQVKLDLTENLTGSKRRRPOTPKPEEAQALDLAGKEL 1352  
Qy 799 DQOEHSAEEGVS-----CDPPATKADSDVDEVRVPE-----830  
Db 1353 FQTPGHTTEAANAAGTKTKMPCSESPPEADPTSTRQPKTPLEKRDVQKELSALKLITQ 1412  
Qy 831 -----NHASKVEGDNTERDLDR--ASKVEPRDEDLVVAQOINNAQRPEDNDSSATC 883  
Db 1413 TSGETHTDKVPGGDKSINAFRETAOKLDL-----AASVTGSKRHRPKKEKAQPL- 1464  
Qy 884 SADEVDCEPERQRMFMDSPKSLNPTGSLVSSPLKPNLDLPLOH-----932  
Db 1465 ---EDLAGKELFQTPVCTDPTTHEKTKTACKS--OPDPVDPTSSKPOSKRSRLKVD 1519  
Qy 933 -----RAAVIPPMVSCTPCNP--IGTPVSGYALQ-----RHIKAMH 968  
Db 1520 VEEEPFALKRTPSAGKAMHTPKPAVSEKNIYAFMGTPVQVKLDITENLTGSKRLQTPK 1579  
Qy 969 ESAL-----LEQORQOEQIDLECRSS-----990  
Db 1580 EKAQALDLAGKELFQTRGHTTEESMTNDKAKVACKSSQDLDKNPASSKRLKTSLGK 1639  
Qy 991 -----TSPCGTSKSPNREWEVLQAPHOLI---TNLPEGVR 1023  
Db 1640 VQVKEELLAVAGKLTQTSGETHTHTEPTGDKS-----MKAFWESPQQLDASAALTSKGR 1695  
Qy 1024 LPTTRTPRPPPLPSSKTTTVASEKSPIMGSGISQGGTGLTSHNOASYTOE--TPRPS 1082  
Db 1696 QLRT-----PKGSEVPEDLAGFI---ELFQ-TP-----SHTKESMTNEKTTAVS 1736  
Qy 1083 -----VGSISLGLPRQOSAKSATLPIYKQEFSPRNSQOPEGLLVRAQHEGVVRG 1134  
Db 1737 YRASQPDLDVDTPTSSKPOPKRSRLKAD-----TEBEFLAFRQTPSAG---KAMH--TPK 1787  
Qy 1135 TAGAIOESITRGTPTSKISVESIPSLRGSITQGTPALPQFG-IPTEALVSGSISRMPLE 1193  
Db 1788 AVGEKIDINTELGTPVQK-----LDQPGNLP-----GSNRL-----1819  
Qy 1194 DSSPEKGRGEEAAGHVIYEGKSGHILSYDNKNAREGTRSPRTAHETSLKRSVESBGN 1253  
Db 1820 -----QTRKEAQ-----ALEELTGFRFLFQTPCTDNPTADEKTKKI---1857  
Qy 1254 IKQGMMSRSPVASPLEGLICRALPRGS---PHSDLKERTVLSGSMOGTPRATESPED 1310

Db 1858 -----LCKSPQSDPAD-----TPTNTKQRPKRSLSLKADVEEFL---AFKRLTPSAGK 1902  
Qy 1311 GLKYPKQIKRSPPIRAFEAGIT-----KGKPYDGTITTIKEMGR 1349  
Db 1903 AMHTPKAAVGEKIDINTFVGTTPVEKLDLGNLPGSKRRPQTPKEKAKALEDLAGKELFQ 1962  
Qy 1350 S---THEIPRODILTOESRKTPVQVSTRIPIEGSISQCTPIKFDNNSGQ-----SAIKHN 1402  
Db 1963 TPGHTEESMTDDKITEVCKSP-----QPDVPKTTSSKQRLKISLGKVG 2007  
Qy 1403 VKSLITGPKSLSRGMPLLEIVPENIKVYVERGYEDVAGETVRSRHTSVVSSGSPSLRST 1462  
Db 2008 VKEEVLVPGKLTQ-----TSGKITQT--HRETAGDKSI--KA 2041  
Qy 1463 LHEAPKAOLSPCIYDDTSAR--RTPVSYQNTMS--RGSPMMNRTSDVTIPPNKSNHERRK 1518  
Db 2042 FKESAKQMLDPAFYGTGMWRPRTPKPEEAQSLDLAGKELFQTPD-----HTEE 2091  
Qy 1519 STLTPQRESIPAKSPVP-GVDPVVS--HSPFDPHHRGSTAGEV-----YWSHLP 1565  
Db 2092 ST-TDDKTTKIACKSPPEESMDTPTSTRRPKTPLGKRDIVEELSALKOLTQTTHTDKVP 2150  
Qy 1566 TOLDPAMPFHR-----ALDPAAAYLFORQLSPTPGYPSQ--YQLYAMENTRQT--ILNDY 1617  
Db 2151 GDEKGINVRETAKOKLDPAASVTGSKRQ-PRTPKGAQPLEDLAGLQELFQTPVCTDK 2209  
Qy 1618 ITSQOMQVNLDPDVARGLSPREOPLGLPY-----PATRGI-----IDLNMPTTI--1662  
Db 2210 PTTHEKTKTI---ACR--SPODPVGTPTIFPKQSKRSLRKADVEEESLALAKRTPSVGK 2264  
Qy 1663 --LVPHPGGTSTPPMDRITYIPGTQITFPPRPYNSASPSGPHPTHLAAAASAEERERER 1720  
Db 2265 AMDTPKAGGDEKMAFMGTPVQKLDLP-----GNLPG-----SKRNPQPK 2307  
Qy 1721 EKERERERIAAASDLYLRPQSEQGRPGSHGY--RSPSPSVRTQETMLQORPSVQGT 1778  
Db 2308 EKAQALDLAGKELFQTPGDKTDEKTTKIACKSPQDPVDTASTKORPK-----2361  
Qy 1779 NGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPA-----SRYNATAADALAAALVDAASAP 1832  
Db 2362 ---RNLKADVEEFLALRKRTPSAGKAMDTPKPAVSEKNIINTFVTPVQKLDLGNLP 2418  
Qy 1833 QMDVSKTKESKHEAARLEENLRS-----RSAAVSPQOOLEQKTLVEKRSVQCLYTSSAF 1887  
Db 2419 --GSKRQPTPKAEALDVGKELFQTPGHTTEESMTDDKITEVSKS-----2466  
Qy 1888 PSGKPOPHS-SVYSEAGKDKGPPPKSYEE-----LRTGKTTITAAINFIDVIITR 1939  
Db 2467 ---QPESFKTSRSKQRLKIPLVKVDKKEEPLAVSKLTTSTGETTQHTB-----PTG 2517  
Qy 1940 QTASDKDAREGSSQSSSSSLSSHR-----YETPSDAIEVIS 1977  
Db 2518 DSKSIKAKESPQKQILDPAASVTGSRROLRTRKEKARALEDLVDFKELFSPAGHTEESMT 2577  
Qy 1978 -----PASSPAP-----POEKLTQYQPEVVKANQAEENDPTROYEGLP 2014  
Db 2578 IDKNKTIKCPKPPPELTDATSTKCPKTRPRKEVKEELSAVERLTQTSQGSTHHTKEPA 2637  
Qy 2015 -----HHYRPOQESPPQOOLPPSSQAEGMQVPRTHRLITLADHTCQIIT 2060  
Db 2638 SGDEGIKVLKORAKKKPNPVEEESRRRPRAPKEKAQPLEDLAGTELSETSGHTEESIT 2697  
Qy 2061 QDFARNQVSSQTPQOPTST-----FQNSPSALYSTPVRTKTSNR--YSPE 2104  
Db 2698 AGKATKIPCESPPPLEVDDTTASTKHLRTRVQKVQVKEEPSAVKFTQTSGETTADKEPA 2757  
Qy 2105 SOAOSVHHORGSVRSPENLVDKSRGSPKSPERSHVSSEYEPISPPQVPPVHEKODS 2164  
Db 2758 GEDGKIKALKESAKQTPAPASVT--GSR--RRPRAPRESAQAIEDLAGFKDPAAGHTEES 2814  
Qy 2165 LLLLSQRGAEPAEQNDARSFGSISYLSFTFKLNTS-----PMVSKKQEIFRKLNS 2218  
Db 2815 M-----TDDKTTKIPCKSS-----PELEDTATSSKRRRPRTRAQVVEKEELLA 2857







Query Match 3.1%; Score 391.5; DB 2; Length 2282;  
Best Local Similarity 18.8%; Pred. No. 1.2e-07;  
Matches 468; Conservative 294; Mismatches 886; Indels 843; Gaps 115;

Qy 2 SSGYPPNQGAFSTEQSRYPHPSVOYTPNTRHQOEFAVDPYRSHLEVSQASQLLQOQQ 61  
Db 33 SSAPYP---GSGTTPASATQELLATOFFGSPSQE-----KTGQ 69

Qy 62 QOOLRRRPSLLSEPHGSDRQERRTS-----YEPFHGSPS----- 97  
Db 70 QOKPARRPSIEASVHI--SQLQHLPLTAPFMPGPKPEHLLGSTWQLVDPMRPGSGSFA 128

Qy 98 PVDHDSLESKRPRLEQVSDSHFORVSAVLP-----LVHPLPEGLRASADAKD 146  
Db 129 PGSHPO-----SQLPSSH-----ASILPEELPGIPKVFVPRPSQVSLKPAEAAH- 174

Qy 147 PAFGKHEAPSP-----ISQPCGDDQNASPSKLSKEELIQS----- 184  
Db 175 -----KERKPOKPKYICQYCSRPC-----AKPSVLQKH---IRSHTGERPYPCGPGFSF 222

Qy 185 -----MDRVDREIAKVEQOILKLKKKQOOL 209  
Db 223 KTKSNLYKHKRSHAHRTKAGLASGSSSEWYPPGLEMERKIPGE--EFEE---PTGESTDS 277

Qy 210 EEEAAKPEPEKPPVPPVEQKHRSIVQIYYD--ENRKAEEAHKIFEGLPKV----- 261  
Db 278 BEETGAASGPSTDVLPKP--KHPLLSLLSYSGSHGSSQERCISLSQSSTGSPLEDPAFP 334

Qy 262 -----ELPLYNQPSDTKYHENIKTNQVMRKLLIFKRRNHARKQROKTCQRYDQIME 316  
Db 335 AEASSEHPLSHKPDETHIKOKALRLSERKKLI---EEOFTLSPGSKGSTESGYFSRSE 391

Qy 317 AWEKKVDRIENNRKAKESKREYKQPEIRKQEQEORFORVQORGAGLSATIA- 375  
Db 392 SAEOQV-----SPPNTNAK-----SYAEIIF-----GCKRGIRGORTSMLASTSTQP 432

Qy 376 -----SEHEIS-----BIIDGLSEQEN-NEKOMRQLSVIPPMFDAPQRRVKFTNM 420  
Db 433 LPLSSEDKPSLVPLSPRTQVIEHITKLITINE-----AVVDTSEIDSVKPRSSLTR 486

Qy 421 NGLMEDPMKYKDFQFNVWTDHKEIFKDK---FIQHPKNFGLIASVLEKSPVD--CV 475  
Db 487 RSSVESP-----KSLYRDSLSHGKTKQEQSLLSLQHPSPSTHPVPLLRSHSPSACT 542

Qy 476 L-----YYV-----LTKKNENKALVRNRYGKR--G 500  
Db 543 ISTHHHTFRGSYFDDHVDADDEVFSRNTPVTSHPMDLKRHAALPLUGGEYSSEEPGS 602

Qy 501 RNOQIARPSQPEKVEEEDKAEKTEK-----KEEEKK----- 533  
Db 603 SKDPTSKPSDEP--EPKESDLTKTKKGFKTKGANYECTICGARYKKRDNYEAHKYYCS 660

Qy 534 -----DEEEKDEKDS-----KENTKEKID----- 555  
Db 661 ELQITKAHSVGAHEVEKTAQEPWQSMHYMKLGLATLELPLRKRKRKESKLGDEEPPAF 720

Qy 556 -----GTAETEEREQATPR-----GRKTANSQGRKGRITRSM 589  
Db 721 ACPGPGSETAHRPLJGTSKPAEASKAPSLSDPRASSGFLPSQLGQNGRRGQCCKKF 780

Qy 590 TNEAAAAASAAAAATEEP-----PPPL----- 611  
Db 781 TVIQTSSFEKSDPEQPSGLLEEDKPPAQFSSPPPPAPHGRSAHLSQLPLVRQPNQIYPEI 840

Qy 612 -----PPPEPISTEPVETSRWTE-----EEMEVAKKGLVEHGRNMAAI 650  
Db 841 LVTEEPDRPDTEPEPPPEKPEKTEEFQWPQRTQTLAQLPAEKLPPKKRL-----RL 892

Qy 651 AKWYGTSEACQKFNFYNYKRRHNDLNLQHKOKTGRKPREERDVQCESVASTVSAQE 710  
Db 893 AEMAQSGGESSFES-----SVPLSRSPSQESSISLS-----GSSRSASF 931

Qy 711 DEDIEASNEEENP--EDSEVEAV-----KPSEDSDPENATSRGTBPA 750  
Db 932 DREDHGKAFAPGPFSDTSTKSLGSHMLTVPSHHPHAREMRRSASEQSPNVPHSSHMTETR 991

Qy 751 VELEPTTTTASTSPSLAVPSTKPAEDSEVETQV-----NDSISAEATAEQMDVDOQEHSA 805  
Db 992 SKSFEDYGLSLP--TGPLAVPAAPPPPAAPPARRKCKFLVRQASLNRPPAEAEAVPRENR 1050

Qy 806 BEGSVCDPPPKATKADSDVDVVRPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQ 865  
Db 1051 AVRPAASKPTKTSKSVPOISVGTTCGGFS--GGKSQMDRPPPLGSS--PPYTEALQVFP 1106

Qy 866 INAQRPESQSDNDSSATCSADEVDGE--PERQRMFPMDSKPSLNLPTGSLVSSP----- 919  
Db 1107 LGTQLPPASLSFLSQLQLLPQEQOSSEFP--TQAMAGLLSSPYSMPPLPPLPSLQFAPPLPL 1165

Qy 920 -----LKNPDLPLQLOHRAAVI-----PPWVSCPTPCNIPIGTPVSGYALYQRIKAMHESA 971  
Db 1166 QPTVLHPSQLHLQPLLPHAADIPFQQPPSFLPMP- -PAPSTLSGY----- 1209

Qy 972 LLEQRQORQEOIDLECRSSTSPCGTSKSPNREWEVLQAPHQLITNLPEGVRLP----- 1025  
Db 1210 FLPLQSQFALQDPEIEHSLPVKTSLPP---LATGPPGPS--STEYSSDIQLPVPVTPQA 1265

Qy 1026 -TTRPTRPPLIPSSKTTVAS-----EKPSF--IMGGSISQ-----GTP-GTY 1065  
Db 1266 TSAPTSAPPLALPACPDAMVSLVVPVRIQTHMPSYGSAMYTLLSILVTQSPGSPASTA 1325

Qy 1066 LTHNQASQYTOET-----PKPS----- 1082  
Db 1326 LTKYEPPSKMTVCEADVYEAEPGPSISKEQNRGYQTYLVRPERKGTSLSSGILLS 1385

Qy 1083 -----VGSISGLPQROESAKSATLPVIKOEES----- 1111  
Db 1386 EGCSTAGSKRVLSPPAGSLMETMETQOQR-----VKEEASKADEKLELVSTCSV 1438

Qy 1112 -----PRSQNSQPE--GLLVRAQHEGVVRGTAGA-----IOEGSITRGT----- 1148  
Db 1439 LTSTEDRKKTEKPHVGQGRSRREAETLSLSDSDVDPKELSPLSHSLSHGAPGSEAL 1498

Qy 1149 -----PISKISVESIPSLRGSITQGTALPQGTPTPEALVKGSIIRMPIDESSPKGREE 1203  
Db 1499 KEYAQPSKKAHRRGLPPM--SVKKEDPK--EQTDLPPLA---PPSSLSDTSKPAKLQ 1551

Qy 1204 AASKGHVIECKSGH-----ILSYDNIKNAREGTR-----SPR-----T 1237  
Db 1552 EGTOSKVLQPLPSLHTTNVSWCYLNYKPNHIQHADRSSVYAGWCISLNPNLPGVST 1611

Qy 1238 AHEISLKRYSVESGNIKQGMRESVPSAPLEGILICALPRGSPHSDLKERTVLGSGIM 1297  
Db 1612 KAALSLLRSKQVS--KETYTMATAP--HPEAGRLVPSNRSKPRMTEVHLPSVVSPESQ 1666

Qy 1298 QGTPRATTE-----SPEDGLKYPQIKRESPIRAPEGAITKGPYDGTITTKEMGRSTH- 1352  
Db 1667 KDPARVEKEEKGKAEET--PTSKRGEPARVKIPEGGYKSNEEY--IYVRGRGRGRYV 1721

Qy 1353 -----EIPRODILTOESRKTPEVVQSTRPII-----EGSISOGTPIKTDNNSSQ 1396  
Db 1722 CEEGIRCKFKPSLKKHIRTHTDV-----RPYVCKHCHFAFKTKGNLTKHMSK----- 1770

Qy 1397 SAIKHNVKSLITGPSKLSRGMPPLEIV-----PENIKVVRGKVEDVKAG-- 1441  
Db 1771 ---AHSKKCOETGV-----LEELEAEBCSTDDLHQDSEGOGEAEVHEHOFSDLESDS 1821

Qy 1442 -----ETVRSRHTSVWSSGSPSLRSTLHEAPKQSLPGIYDDTS---ARRTPVSYQ 1489  
Db 1822 DSDLEDEDEEEEEEESODELSGP-----CSEAAPCLPPTLOENSSPVEGPQAPDSTS 1875

Qy 1490 NTRMGRGSPM---MMRTSDVTTPPNKSTNHKRKSLTPTQRESIPAKSPVPGVDVWVSHSP 1546  
Db 1876 DEVPEGGSISBATHLTASSCSTPSRGTOGLPRLGLAPLEKMDKMSAPSP--KATSP---RRP 1931

Qy 1547 FDPHHRGSTAGEVYWSHLPTQLDPAMPHPHALDPAANAALFQORQLSPPTPGYPSQVLYAM 1606











Db 1164 PQTLDSTQIIIEPTTVVGVPIVEEVTQTTSDQAEYPPSPVTSFQSL-DLELTISP 1222  
Qy 1248 ESVEGNIQOGMSRESVSALEGICRALPRGSPHSDLKERTVLS-GSINQGTPTATTE 1306  
Db 1223 EATRESYHPSLQQTIVNPPPEHLVTHSEQVTHQHPNLTEATVQPLDPLDITIPQPTTE 1282  
Qy 1307 SFEDGLKPKQIKRES-----PIRAF-----EGAITKGKPYD-GIT 1342  
Db 1283 G-----ELPQLDSTQIIIEPIKVVVALPVVQVQSODAEYTTSTVSFQPLDQELT 1337  
Qy 1343 TIKEMGRSHE--IPRODIL-----TQSRKTPPEVVQ-----STRPII 1378  
Db 1338 ITSEAIRPHPTVPQQTITVHTKPLVLIHSEQTHQHPNTEVTVPQLDLELTMTPTQTA 1397  
Qy 1379 EGSISOGTPIKFDNNSGSAIKHNVKSLITGPKSLRGMPLIIVPENIKVVERGKVEDV 1438  
Db 1398 EGELPQ-----TQDSTTQIIIEPTTV-----VGVPI-----YEEV 1429  
Qy 1439 KAGETVRSRHTSVVSGSPVLSRLTIEAPKAQLSPGIYDDTSARRTPVSYQNTMSRGS-- 1496  
Db 1430 TVQTT--SQDQAEYPPSPVTSFQSL-----DLGLTITPEPTEHFIQTQTV 1473  
Qy 1497 --PMNRTSDVTIPNKNSTNHERKSTLTPTORESIPAKSPVGVVDVPPVSHSPDPHHRGS 1554  
Db 1474 FPMY--TDVTLQQVSVQHLK-----PTEGIVQPLDLELTITP-QPTPEGE 1517  
Qy 1555 TAGEVYNSHLPTQDPAHPFHRLDPAALVLFQRLSPGTPGSPYOXYLYAMENTROTIL 1614  
Db 1518 LSQIVESTTQNK-EP-----HK--EVAPVPVYQAVVPT--FSQYQAEYKSLQPLDL 1567  
Qy 1615 NDYITSOQOVNLRPDVAR---GLSPREQLGLPYPATRGI--IDLTMPTTILVPHPGG 1669  
Db 1568 ELTVTSEPTKAYHSTISKSLAINPQVHIQHPNPAEATVQPLDLELTISSSLQPTAEG 1627  
Qy 1670 -----TSTPPMDRITYITPGQITPPRPYNSASPSGPHLAAASAEERERE 1717  
Db 1628 ELLYSMQETVTQISEPPKQVVPVPEQEVAPVAPVQDAKYP--LSISVLSNLDQELT 1685  
Qy 1718 RERKERERERIAAASDLYLRGSEOPRGSGHYRSPSPSVRTO---ETMLQORPS 1773  
Db 1686 LSELLGEAHLQITPDETMVLPKDRQGIYFDHDKHLNLTETVNPQPHLEHVQHOPT 1745  
Qy 1774 V-----FOGNTGTSVITPLDPTAQLRIMPL----- 1799  
Db 1746 IEERSOSIQKTKTQITEPGKVVPLAQSESEVATPMPILKETAPPHPHSMALQSLDEKL 1805  
Qy 1800 ---AGGPSISQ---GLPASRYNTAADALAAVDAASAPQMDV----- 1836  
Db 1806 TIHSHSPGWTQOAHNLKESKGHTTG---KILLDYA--EPNMEIELKHGLFLKLTTEATT 1860  
Qy 1837 -----SKTKESKHEAARLEENLRSAVSEQQOLEQKLEVEKRSVOCLYTSSAPPSG 1890  
Db 1861 ESENTNQMTKSLKQVLTALFTQNKKSLMPALVSESDSQPPNMSLQPLDQELTSSQPHG 1920  
Qy 1891 ----KPOPHSSVYSEAKDKGP---PPKSRYEBELTRGKTTITTAANFIDVIITROIAS 1943  
Db 1921 WPHLPNTPEKIYLHYAEPPTGPFVEPELFF--LKT-----TKSRPVQGTATQMAAS 1971  
Qy 1944 DKDARERGSSQSDSSLSLSHRYETPDSADAEIVSPASAPPQEKLOTYQPEVVKANQAE 2003  
Db 1972 PKEMVSRAPENKEAVLS-----GPGE-----DQDESPPNMSLQSLDQELTSSQPH 2019  
Qy 2004 ----NDPTROVEGLPHVRP----- 2019  
Db 2020 GWIPHPNTHGKIYLHYAEPPTGPFVEPPDLFLTKTKSPKVEVWTLRTDKSRKEMVSQS 2079  
Qy 2020 -----QOESPSP-----QQOLPSSQAEG-MQGVPRTHRLITLADHIC 2056  
Db 2080 KYEEAVLPVHGEGEESRSPNMSLQSLQELTSSQPHGWPVHPNTHGKIYL--HYA 2137  
Qy 2057 QIITQDFARNQVSSQTQPOPTSTFQNSPSALV-----STPVR--TKTSNRYSPESQASV 2110  
Db 2138 -----EPPTGPFVEPPDLFLTKTKSPVQGTATRMVKSPEEMVS-- 2177

Qy 2111 HHQRPGRSVSPEN--LYDKSRGSRPGKSPERSH-----VSSEPEYPI-SPQVP 2156  
Db 2178 -----LDPENKEAVPPAQEGKGESPPSNMSLQSLDHELFMSSQPHGWIPHPKPTP 2229  
Qy 2157 VWHEKQDLSLLSQRGAEPARNDARSQGISYLPSPFF-----TKLE-----NTSPMVKSK 2208  
Db 2230 -----DKIYL--HYAEP-----PTGPFVEPPDLFLRTTKSKPVOGTTTEMAKSP 2272  
Qy 2209 KOEI-----FRKLNSG--GGSDMAAQAOPGTEIFNLPAVTTSGSVSRGHS-FADPASNL 2261  
Db 2273 KEMVSOTPEYKEAVLSGPGEDQESPNT--SLASLDQEVAMSSQPHSGVPHPKPTP 2329  
Qy 2262 GLEDIIRKALMGSDDDKVEDGHVVMVSOPMG--VVPGTANTSVVTSETRREBGPDS----- 2315  
Db 2330 G-----KIYLSI--EPPPGFVAPTDLILVKTTKSKPAEWTPRIDKL 2372  
Qy 2316 -----PHSGGYCKPKLISKNSRKSQPIQGGYLGTERPSSV-----SSVHS 2358  
Db 2373 LKEMVPHS-----PEYEAVPAGHGEQDESGSPNMPQLDQELTSSQPHG 2421  
Qy 2359 EGDYHROTGP--WAWEDRPSSTGTSTOFFNPLTWMLSTSP-----PTPIACAPSAVNOA 2411  
Db 2422 WVPHPNTPGKIYLHYAEP--TGPFVEPPDLFLRTTKSKPVOGSKETAKSPKEMVSQ 2479  
Qy 2412 APHOONRI-----WEREPAP 2426  
Db 2480 TPEYKEAVLSGPGEDQDESFP 2501  
  
RESULT 23  
T16871  
hypothetical protein T13H2.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C;Accession: T16871  
R;Wu, X.  
submitted to the EMBL Data Library, October 1995  
A;Description: The sequence of C. elegans cosmid T13H2.  
A;Reference number: Z18593  
A;Accession: T16871  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Residues: 1-2215 <WUX>  
A;Molecule type: DNA  
A;Cross-references: EMBL:U39653; NID:g1049397; PID:g1049401; PIDN:AAB52495.1; GSPDB:G  
A;Experimental source: strain Bristol N2; clone T13H2  
C;Genetics:  
A;Gene: CESP:T13H2.4  
A;Map position: X  
A;Introns: 112/2; 136/1; 167/1; 196/1; 649/1; 757/1; 850/1; 882/1; 985/1; 1046  
  
Query Match 3.0%; Score 381.5; DB 2; Length 2215;  
Best Local Similarity 18.7%; Pred. No. 2.8e-07;  
Matches 444; Conservative 324; Mismatches 862; Indels 741; Gaps 105;  
  
Qy 165 CGDONASPKLSKEELIQSMDRVDRITAKVEQOI-LKLKKQOOLKEEAAPPEPEKPV 223  
Db 112 CGHEPNL---RMNQE---DSSVVADESAGSPEIGLSLOLKVEVEDSMSPQLRNWEE 165  
Qy 224 SPPEVQKHRSIVQIIYD---ENRKA-----EEAHKIFEGLPKRVLPFLYNQPSDTK 273  
Db 166 TPPTRRSTKMDLLIGDVKPPGRIRATSYSCDIKRSGLISAHFASNDTVCVWVDQTS 225  
Qy 274 VYHENITNQVMRKLLIFPKRRNHARKQREKTCQYDQLMEAWKKVDRINNPRKA 333  
Db 226 IV---VASNOSTRDAL-----ASLCOR-----DSKVERMETAARISW 259  
Qy 334 KESKTREYVEKQPEIRKOREQOORFORVORGAGLSATIAARSEHEISEIIDGLSEQENN 393  
Db 260 KLEDRVTEKTTURLVHEEYANARIOQNNRNTQSLCSQPKRSRRVOGFD----- 311  
Qy 394 EKQMRQLSVIPPMFADAEQRVRVFIN-MNGLMEDPMKVYKDRQFMVWYVTOHEKEI--FKD 450



Db	312	-----PTNLTAEEQAQAFYTYLNTYQODFGKTI-----GNLLHSEHKOLOQPLFN	354
Qy	451	KFIQHPKNEGLIASYLERKSPDCVLYLYLTKKN---ENYKALYRRNVYKRRGRNQOIAR	507
Db	355	K-----LHTSSVP-----LHRLPQITFVN---STVRLEYSSSES-----	387
Qy	508	PSOEKVEEKEEDKAEKTEKKEEKKDBE-----EKDEKEDSKENTKEKDKIDG	556
Db	388	PSNNCCDAQNDRGHFGMSRPLENNTEHNATGISTSFLEKDSAPNPEDST-----	439
Qy	557	TABETTEREQATPRGRKKTANSORGRKRITRSMTNEAAAAAASAAAAATEPEPPPLPPPE	616
Db	440	--BSVLNREDASCTTQKTRNS-----FEECVYNPALVLAAVNRVST-----PAT	481
Qy	617	PISTE---PVERTSRWTEEBEYAKGLVEHGRNMAAIKAVGCTKSEACKNFYFNKRRH	673
Db	482	PVINDVGPPVVS-----ETANK-----AVVSIEGSMKIEOC-----H	514
Qy	674	NLDNLLQHQHKTSRK---PREERDVSOCEASVASTSAQEDEDIA-----SNE	719
Db	515	GPSNPQOLNPNETTRKSVSPLO-----RSMSTPASEATDEDLIAFLREPLALARAEE	567
Qy	720	EENPDESEVEAVKPSDSDENATSRGNTAPAVEL-----EPTTETAPSTSPSLAVPSTKP	774
Db	568	QEALRKKAEMPKLKRKRAKNSAKRPVOTDAVETVKLEKPKAQSKSKNSKRSKSTPS---	624
Qy	775	AEDSEVETQVNDISAEATAEQMDVQOQBSHAEEGSCVDCPPATKADSVDVVERVPENHAS	834
Db	625	-----SFANDSSLSLEPLEM---ROHFEGSELQKPSPELH-----PGTQGS	663
Qy	835	KVBGDNTKERDLDRASEKVEPRDEDLVAAQINAOQP-----EPOSNDSSATC	883
Db	664	SLRNFIRLDWPLQALQEV-PLG---AVAGQLRVQKALYKMDGMNGIAPESASTSQTF	719
Qy	884	SABEDVDGPERQRMFPMDSKPSLLNPT-----GSI-LVSSPLKPNPLDLPLOHRA	934
Db	720	MDPAPPLPPKK-----AAPKRKKPTKAETEAAGNVSVVEKPKPRAPSKKKAQGV	772
Qy	935	AVIPPMVWCT-----PCNIPICTPVSGYALYQRIKAMHESALL	973
Db	773	QQEQHQMISMTSAHQGTPEHPQSHOSTHOPLPVSANLLPLITGY-----	816
Qy	974	EEQRQEQEIDLE-----CRSSTSPCGTSKSPNREWEVLQ	1009
Db	817	QQOQQOQQPPSQOHHGVVNNNTTNGYVSDNGSGSSSSSSSVYSYSYMQPTSDPP	876
Qy	1010	APHOLITNPEGVRLPTTR-----TRPPPLIPSSKTT--	1043
Db	877	KKHQSASMLTPTTFGSSQPEATNGSSQIVTDKNQYHLYNDNNTYHPQAMYSNQTSN	936
Qy	1044	-----VASEK-----PSFIMGSSISQGTPTYLTSHNQASVTOETPKPSVGSISLG	1089
Db	937	GYGGYEDVNNQFOQPDYPLPSVESQVSQBSQESNITYHSSMPVTPISQOANNGSYNP	996
Qy	1090	LPRQESAKSATLPYIKQ-----EEFSR-----SONSQPEG-----LLVR	1125
Db	997	MPYTKDNRESNSSHLOQDHEDLSVEQEAQYAPSDNFFEMQNSRDSQETNSAPLSMAR	1056
Qy	1126	AQHEGVVRGTAGAIQBG-----SITRGTPTSKISVESIPSLRSGISITQGTALPQTGIPTEAL	1182
Db	1057	ASNASPFDELGYLDSGPSTSTQDDPFADIELSDPPL---SI---TPQVAQEPVQT-TI	1109
Qy	1183	VKGSISRMPIEDSSPEKGREEAASKGHVYBCKSCHILLSYDNKNAREGTRSPRTAHEIS	1242
Db	1110	MSNRVTKTPSSETHPO-----SYANQYVPMNPSLPVSSHSES	1147
Qy	1243	LKRSYVESGNTIKQ-----GMSRESVPVS-----	1266
Db	1148	QVSTKLLPSNFKEDDIRHSTCPDNCNKCVSKEKSNADISMQPSASTIVTVPQSQORDQH	1207
Qy	1267	-----APLEGLIC---RALPRGSPHSDL---KERTVLSGSIQMGTPRATTESFED	1310

D	b	1208	LEOQGVQEUAATDDBOICIVVYKYLPHETISHTNLGRRPQQSKPSGVGVNGTSHGLTNQGPP	1367
Q	y	1311	GLKYPQIKRESPIRAFE---GAITKGKPYDGITTIKEMGRSIEHPRQDILTQESRKT	1367
D	b	1268	PLPPPKESSAKPKKNAFEDMNSSGHAY-----	1303
Q	y	1368	PEVVOSTRIPIEGSISQ-----GTPKFDNNQSQAISKUN--KSLLTGPSKLSRGMP	1418
D	b	1304	PEV-----SIGKLKQOHLRNPNQFNNISOSQIDLATLQOVSMLSP-KLQRYSE	1352
Q	y	1419	PLET-----VPENIKVVERG--KYEDVKAGETVRSRHTSVSSGPSVLRSLTHE	1465
D	b	1353	POTMKPFHHQKAQVHEKKNNHQSGAKAPEIQRHYNLVNQVPTQOSSGPA-HQSTIH	1411
Q	y	1466	AP-KAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNRTSDVTIP-----PNKST-	1513
D	b	1412	APMKQIK--LTSLOSNKTSQEHQSTTLQ-APVEEAPS-QPFENFLEQDSPDKVTE	1466
Q	y	1514	-NIERKSTL-----TPQBESIPAKS-PVPGVDPVSSHSPDPHHRGSTAGEV	1559
D	b	1467	INNVA SVLSSPD EFRGQKFTKTEKHLDDGT SKFTP-----KKQSHVDRRSRGSF--V	1518
Q	y	1560	YWSHLPOLDPAMPFFURALDPAAAYLFORQLSPPTGYPS---QYOLYAMENTRQTILN	1615
D	b	1519	HYVGROQMISOQONRN-----AYVOQOOQP NASSLSTMNEQYEV-VLDNTN--LQ	1567
Q	y	1616	DYITSOOMOVNLRPDAVARGLSPREO--PLGLVPYATRGIIDLITNMPTTILVP---HP----	1667
D	b	1568	NPMSEGOVMNYQTSIVOOTSVEEQGPLQL-----QNOIQVTNQ-QTHRVQYQHHPVQHN	1621
Q	y	1668	--GGISTPMDRITYIPCTQITFFPRPVNSASMSGHPTHL---AAAAEAERERER	1720
D	b	1622	QRNGPTKAAPKR-----TPKPAPVQSRSAVALHERAQMI VDFAKTOPADEQDTVQ	1672
Q	y	1721	EKERERTIAAASDLRLRPGSEGCPRGSGHGYSRPSPSVRTQETMLQRPSPVGTNG	1780
D	b	1673	LQDHEHQ-----YNSQQP-----EYQHQLVMQEQNESQEQSQVFQHQH-	1712
Q	y	1781	TSVITPLDPTAQLRIMPLPAGSGSTSOGLPASRYNTAADALAALVDAASAAPQMDVSKTK	1840
D	b	1713	-----HQAOQELAEENG-----SMGMTQOOCROCSOLOQLOPPO-OOMIOIOOY	1759
Q	y	1841	ESKHEAARLEENI---RSRSAAVSEQOOLEQKLTLEV-EKRSVQCCLYTSSAPPSPGKPOHS	1896
D	b	1760	QNVGP SHIQ NVAP RVQV QOQ PMYA QOQ OQ PMMA QOORS MQE YQV QOG VQT GATH T	1819
Q	y	1897	SVVYSEAGKDGP PPKSRYYEE---LRTGKTTITAA---NFIDVIITRQIASDKRARER	1950
D	b	1820	VLMTS-----PPMHNHESQANVLRVAPKQTVIKVSADPKPEDEAKILK ILKDKLASDD	1871
Q	y	1951	GSOSSDSSSLSS-SHRYEF-----PSDAIEVISPASSAPPEKLOTY-----	1992
D	b	1872	GONAPKSPVLRVKHIREFINIIAULPSLTIEELFDKADTPPDEYMEEYMDFKDKITHL	1931
Q	y	1993	--OPEVVKANOANDPTRQYECPLH-----HYRPOQESPQ-----	2027
D	b	1932	MKDEPCVPEINVEDVLSTPLLEPALHGRPKRLAIAREGTRTVTTFKKRKRTASGPEGDAK	1991
Q	y	2028	-----QQLPSSQAEG-MGOVPRTHRLLITLADHICQIITQDFARNQVSSQTPQ	2074
D	b	1992	KRKPTNNHTNMHTPDASPACTSSGSSSS-----GMOSLSLSTTCGMSSSS	2040
Q	y	2075	QPPTSTFONSALVSTPVTRKTSNRYSPESOAQSVHHORPGSRVSPENLVDXKSGSRPG	2134
D	b	2041	SPASDEF-----ISTHLITPQEHSSPLTPTPIINHEYVQOYSSPVDIFSQSTSEPG	2094
Q	y	2135	KSPERSHVSEPEYPTSPQPVVHKQDSL	2165
D	b	2095	-----PSSRPIRTGVHIREAL	2111



[illegible]



Db 2173 LLPSTATVRAADGMISTIN-TPIAATLP-----ITTPASVLRPMVRGG---MYRPG 2222  
Qy 1502 TSDYTIPNKTNNRSTLTPTQRESIPA-----KSPVGVDPVYVSHSPDFPHRGST 1555  
Db 2223 SGGVAVPLTSL-----TRVPMIAPRVLGAGLYRYPAPSRPIASTIP-----PA 2269  
Qy 1556 AGEVYWSHLPTQLDPAMPFHRALDPAAAYLFQRLSPTCPYPSOYQLYAMENTRQTLN 1615  
Db 2270 EGPVYLG-----KASGAGPP-----2290  
Qy 1616 DYITSQOMVNLDPVARGLSRPOPLGLYPATRGIDLTNMPPTILVPHPGTSTPPM 1675  
Db 2291 -----RDELPAQ-CAREEPLS-----TTAPPVRIKEAPVAQAPAP- 2324  
Qy 1676 DRITYIPTQITFP--RPNYSASMP-----GHPTLAAASAE-RER-----ER 1718  
Db 2325 -----PPGOKPAGDAAGSGVGLGRPVMEKEASQEDRQKQOQLLOLER 2371  
Qy 1719 ER-----EKERERERIA-----AASDLYLRP-----GSEQGRPGS 1750  
Db 2372 ERVELEKLRLQLEELERERVELQRHREBQLLVQRELOELQTIKHVLLQOQOERQAQ 2431  
Qy 1751 HGYVRSPPSVRTQETMLQQRPSVFQGTNGTSVTPDPTLAQLRIMPLPAGGPSISQGLP 1810  
Db 2432 FALQRELAQORLQLEQTQLOQLOQ-----QLEEQKQKQKAPFPATCEAPSRGPP 2483  
Qy 1811 ASRYNTAADA-----LAALVDAASAPQMDVSKTESKHEAARLEENLRSRAVSE 1862  
Db 2484 PAATELAQNGYWPPLTHPIAVAGTEGP-----QOAREPV-----LHRLGPSASDMSL 2534  
Qy 1863 Q--QOLEQKLTLEVKRSVQCLYTSAPFSKQPQPHS-----SVVYSEAGKDKGPPPKSR 1914  
Db 2535 QTEOWEAGRSIGIKRHSMPRLRACPEPGDPSTVRRITADSSVQTDDEGEG-----R 2589  
Qy 1915 YEEELTRGKTTIYANFIDVIITQIASDKD-----ARERGSSDSSSSLSH 1964  
Db 2590 Y-----LLTRRRRTRSA---DCSVQ---TDBEDNAEWEPQVRRRRSRLSRHSDSGSD-SKH 2639  
Qy 1965 RYETPSDA-----IEVSPASSPAPPOEKIQTQYQ-----EVKAN 2000  
Db 2640 EASASSAAAAAARMSVGIQTLSDCSVQTEP-EQLPRVSPATHIAATDPKVEIYRI 2698  
Qy 2001 QA-----ENDPTRQYEG-----PLHHRP-----QOESPS 2025  
Db 2699 SAPEKTGRGESLAQTEPDGQAQVAGPQLIGTAISPYPGLGIVTPGALGREKKKPD 2758  
Qy 2026 P-----QQOLPPSSQAEGMG-QVPRTHRLI-----TLADHICQIITQDFARNQ-----2067  
Db 2759 PLEIGYQAHLPPELSQVRSQPKSPQVLYSPVSPISPH--RLLDTSFASERLKAHV 2816  
Qy 2068 -----VSSOTPOQP-----PTSTFQNSPALSVPVTKTSNRYSPESQAGSVHHRP-G 2116  
Db 2817 SPQKQFTADSTLRQTLPRPKTLQRS-----LSDPKLSPTAESAKERSLYHQGGGLG 2872  
Qy 2117 SRVS-----PENLVKRSRSPKSPERSH-----VSSEPYEP-----ISPQVPPVYHE 2160  
Db 2873 SOVSALPENGVLVRKVTLPSPPEEAHLPLAGQVPSQVLAASLLQRLGAGTTPVAT-- 2930  
Qy 2161 KQDSILLLSQGAEPABORNDARSPGISYLPSTFTKLENTSPMVKSKQBIFRKLNSG 2220  
Db 2931 -KASLLRELDRLVREHES-----TKLRKQABDEEKEBIDAKLYLE 2974  
Qy 2221 GQSDMAAAGPTEIFNLPAVTTSGSVSSRGHSFA-----DPASNGLIEDIIRKALM 2272  
Db 2975 LG-----ITQKESLAKDRVGRDYPPLRGLGEH-----RDLYS 3007  
Qy 2273 GSFDDKVEDHGVMSQPMGV---VPGTANFSVTSGETRREE-----GDPSPH 2317  
Db 3008 DSELNQLRLOG--CTTAGQYVDYPASAANPATPSGTAFQCPFRFPAAQTYTAGSGPT 3065  
Qy 2318 SGGVCKPKLISKSNRSKSPi--PGOGYLGTERPSSVSVHSGDYHROTP-----2367

Db 3066 QNGFLAQHQAPTYTGPSTVPAPYTPPGTSYPAEPGLPSQPAFPHGTCHYAAPTPMPTQSAP 3125  
Qy 2368 -----GWAWEDRSSSTGSGTOFFYNPLTMRMLSTPTPTPIACAPSANVQAAPHQONRIWE 2421  
Db 3126 FVQADSHAAHQKPRQSLADLEQKQVPTNYEIVISSPAVTVSTSTSETGYS-----3175  
Qy 2422 REPAPLLSAQYE 2433  
Db 3176 --GPAVSSSYE 3184  
RESULT 25  
T16870  
hypothetical protein T13H2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Sep-2000  
C:Accession: T16870  
R;Wu, X.  
submitted to the EMBL Data Library, October 1995  
A:Description: The sequence of C. elegans cosmid T13H2.  
A:Reference number: Z18593  
A:Accession: T16870  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2447 <WUX>  
A:Cross-references: EMBL:U39653; NID:g1049397; PID:g1049400; PIDN:AAB52494.1; GSPDB:G  
A:Experimental source: strain Bristol N2; clone T13H2  
C:Genetics:  
A:Gene: CESP:T13H2.3  
A:Map position: X  
A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1  
C:Superfamily: RING finger homology  
F:158-207/Domain: RING finger homology <RRN>

Query Match 3.0%; Score 376.5; DB 2; Length 2447;  
Best Local Similarity 19.1%; Pred. No. 4.9e-07;  
Matches 492; Conservative 309; Mismatches 892; Indels 881; Gaps 119;

Qy 98 PVDHDSLESKRPRLEQVSDSHFQVRSAAVLPLVHPLPEGLRASADAKDPAGGKHEAPS 157  
Db 199 PTCRNLGKKR---ELOQDPDFDLIYQVVSRSIVGRMAENREHEKDYVYFGRKYIEG 255  
Qy 158 SPISQPCGDDONASPSKLSKEELIQSMR-----VDREIAKVEQOILKLLKQKQ 207  
Db 256 GSDNKKRYGIDPN---SKLKAPRLKSAKKIRWFHESDEGSGVRKVM-----SKGA 307  
Qy 208 QLEEE-----AAKPEPEKPVSPPPVQKHRSIVQI-IYDENRKKABEAHKI 253  
Db 308 PKEDDTNVLNDKEGTSVAEKEVLEEGEMDFPIEKSSDDEQTDLDDEESMLDSPEI 367  
Qy 254 FEGLGPKVELP-----LYNQPSDTKVYHENIKTNQVMRKLILFFKRRNHAR-----300  
Db 368 SD--NEDVSKPSCSTSKTTNRSDESSESDMSDRLNELQKKRKM--KRWKVPKTDGSDV 423  
Qy 301 -----KQEQKICQR-----YDQLEAMEWKKVD 323  
Db 424 SNESFDEDSASVATLILKESKKKPCGRPKKFAPELIEGDIPTPSEDSLTSSDEEDD 483  
Qy 324 -----RTENPRRKAKESKTYEYKQFPEIRKQREOER-FORVGORGAGLS 370  
Db 484 NAADPYAFVQKEPNRDPDRDGHPEKDLY---NFDEMDNMNHQVDRKFEKDE-----534  
Qy 371 ATARSHEILSEIDGLSEQENNEKQRLSVIPPMFDAQBRVKFINNGLMEDPMKV 430  
Db 535 -----IHWISD--DSNEHESDEADRESSI-----DSE-----561  
Qy 431 YKDRQFMNVWTDHEKEIFKDKFIQHPRKNFGLIASYLERKSV-----DCVLYYLTKK 483  
Db 562 -----HKEI--SKFLSH-----RQPLNPTSVDCCQV-ITVVKK 594  
Qy 484 NENYKALVRRNYGRRGRNOOIARPSQEEKVEEKDAEKTEKEEKDEEKDEKED 543



595	DVQSAIT-----SKGETS-----PDSSKIE-----EKPDKVSEVSDDEMPHIT	638
544	SKENT-----KEDKIDG-----TABETEERQATPRGKTAN-----SQGR	580
639	ADGTDOTFLNIMEHDEMYGGVLFPRPGDTGISRPKVQAPCTNRLSMNVCPAVLKEGK	698
581	RK-----GRITSMWTEAAAAASAAAAATEEPPLPP	613
699	KKLIVIPPEDYEITSSDETVTLSDEETSPSAEMEQSETSEAGPSTIIKTSGTTEROTGSSS	758
614	PPPISTPEVETR---WTEEMEVAKKGLVB-----HGRNWAIAKMWGTSK	658
759	PSEP-----STSRDRKMKRLDTRRRKLADDSLSDFDVFSIDGNELVATGPI	808
659	EAQCKNPFYKRRHNLNLLQHQKQTSRKPREERDVSQCESVAST-----	705
809	-IKHKVFDYSANRMPKSLDFTGRNAREIPMEETISRLAEQVAHEEYKTHRRQVVLE	867
706	-----VSAQEDEDIASNEENPEDSEVEAVKPSDESPENATSGNTEPAVEL	753
868	AVEAASKKLVVYDITTEEBEIE---EETPEEVYKVASP-----TAPIATE	911
754	EPTTETAPST-----SPSLAVDSTKPADESV-----ETQ-VNDSI	788
912	NPTTSTAPPEEGVAMKETPIEIEFPDPEPCSSAAQARELIIERVKGEQIIEDSLEQN	971
789	---SAETAQOMDV-DOEHSAEGSVCDPPPKATKADSVVDVEVRYPENHASKVEGDNT--K	842
972	RKPSSTTVKSESREQAEPRIEKKDEMEBQOKKADNPTVEVD-KESEASSESDKDFE	1030
843	EROLDRASE--KVEPRDEDLVVAQOIN-----	867
1031	DETLDAQSKTVKISLKHETVSDIELEDFTKGEFVATADAKMIKRTIGEYVSTEFKL	1090
868	-AQRPEQSDNSATCSADENDVGEPEQRMPMDSKPSLLNPTGSTLVSSPLKPNPLD	926
1091	VAQOPAVTDEVLAGFCVRNTDQEFSTIKE-----TGKRTKNPDDESVMV--NFRE	1140
927	LPQLOHRAAIVPMVSCTCPNIPGTPVSGYALYORHI---KAMESALLFEORQOEI	983
1141	SFAAKR-----PVPKRLTNTIERMYIERAHVMVKYKHVDMEPLHMKILIALQKQ--QI	1192
984	DLECRSTSPCGTSKSPNREWEVLOPAH-----QLITNL--PEGVRLPTTRTPRPPPLI	1037
1193	AATCAMLSPV-----TVTPEEHAEQVQLLHLNLPNSILR-----PLL	1230
1038	PSSKTTVASEKPSFIMGGSISQGTPTVLTSHNOASYTQETPKPSVGSISGLPQOESA	1097
1231	-----NNPOFAL-----TLHKAQOQAIQOORAQQAQTKELA-ARQAEA	1270
1098	KSATL--PYIKOEFSFR-----SONSOPEGLLV-----RAQHEGVYRGTAGAIOE	1141
1271	RVEELARKRIAQEDAELKARQKGEQMSNVGIPVSSDQNAQSSNAQQTGLIENQTITNS	1330
1142	GSITRGTPSKTS-----VESIPLRGSTQGTTPALPQTGITEALVKGSGISMP	1191
1331	DSLTRPTNLADNSHLGSESQOIPVIESIQSSSTSEALKESENTKOMPILTPASTVSSKSSAP	1390
1192	I--EDSSPEKGREAAKSHVIYE--GKSGHTLSYDNINKNAREGTRSPRTAHEISLRSY	1247
1391	ATRRSPRCSYDRPSPSPVIRERLGSOGALIN-----RPNRC-----	1430
1248	ESVEGNIKQGMRESPPVS-APLEGICRALPRGSPHSDLKERVLSGSIQMGTPRATTE	1306
1431	-----NIDKRS--RSPISRAPVETV-----RINDHG--QNETILAGNI-----TH	1467
1307	SFEDGLKYPKQIKRESPPIRAFEGAITKGKPYDGTITIKMGRSITHEIPRODILTOESRK	1366
1468	TVE-----TTILEGTSI-----QDOSTI	1486
1367	TPEVOSTRPIIEGTSISQGTPIKFDNN--SGOSAIKHNKYSLITGPSKLSRGW-----	1417
1487	RYDGECSQTTQYIDKTD-----LONSKNGTNVDBEQSNVLKRENDLNLREMLYANRYH	1540

RESULT 26

RESULT  
T02345

102343 hypothetical protein KIAA0324 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C;Accession: T02345

R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D



re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.  
submitted to the EMBL Data Library, March 1998  
A:Description: Sequencing of human chromosome 16p13.3.  
A:Reference number: 214664  
A:Accession: T02345  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1791 <RIC>  
A:Cross-references: EMBL:AC004493; NID:g2996648; PID:AA08453.1; PID:g2996650  
C:Genetics:  
A:Map position: 16  
A:Introns: 1610/2; 1706/2  
A:Note: KIAA0324

Query Match 3.08; Score 376; DB 2; Length 1791;  
Best Local Similarity 19.98; Pred. No. 3.5e-07;  
Matches 396; Conservative 222; Mismatches 803; Indels 566; Gaps 75;  
Qy 681 QHKQKSKPREEDVSOQESVASTVSAQE-----DEDIASNEEN 722  
Db QKSKDSLQSCGSLCAGKSTPPGESYFGVSSLLQKQSQTSDFHRSDTSSPEVR 109  
Qy 723 PEDSEVAVK-PSEDSNATSRNTPEAVELETTETAP-----STPSL--- 767  
Db QSHSESPLOSQKSTSPKGGRSRSS-----PVTELASRPIRQDRGEFSASPMKLSG 162  
Qy 768 AVPTKPAEDSVETQVNDSTIS-----AETAQMDVDQEHSAEAGSVCDPPATK 818  
Db MSPQSRFQSDSSPYVDNSLLGQSLQSLQSLQSLQSLQSLQSLQSLQSLQSLQSL 218  
Qy 819 ADSVDVVRVVRVHASKVEGDNTERDLDRASEKVEPRDELDVVAQINAQRPQSDND 878  
Db FSPFPVQDR-PE--SSLVFKDTRLTPPRRSGA-----GSSPETKEQNS 259  
Qy 879 SSATCSADEVDGPEPQRMPKMSKSLNPTGSL--VSSPLK-----PNPLDLQLQH 932  
Db ALPTSSODEELMEYVEKSE-----EPAGQILSHLSSELKEMSTNFESSPEVEE 308  
Qy 933 RAAV-----IPMWVCTPCNIPIGTPVSGYALYQRIKAMHESALLEE 975  
Db RPVSLTLQSQSOASLEAVEVPSMASW-----GGP-----HFSPEHRELNSP 353  
Qy 976 QRQEQEIDLCEKSTSCGTSPKSPNREWEYIQ-----PAPHQLTNLPEGVRLPTTR--- 1028  
Db LRENSFGSPLEFRNS-GPLGTEMTGFSSEVKEDLNGPFLNQLETDPSLDMKEQSTRSG 412  
Qy 1029 -----PTRPPPLIPSSKTTVASEKPSFIMG-----GSI 1057  
Db HSSSELSPDAVEKAGMSNOSISSPVLDAVPRTPSRSSASSPEMKDGLPRTPSRRSR 472  
Qy 1058 SQGTPGTVLTSHNOAKYTOETPKPSVSGISLGLPQOESAKSATLYIKQEFSPRSON- 1116  
Db SCSPFGLRDGSGTSPRSHLSGSSPGMKDIPTPSRGRSECDSSPEPKALPOTPRRGRSP 532  
Qy 1117 SQPE--GLLVRAQHEGVVGTAGIAEGSITRGTPTSKISVIESIPSLRGSTIGTTPALPQ 1174  
Db SSPENLNKLTPOREK--SGSESSVDQKTVAR-TPLQRS-----RSGSSQELDVKPS 582  
Qy 1175 TGITEALVKGSI-----SRMPEDSSPEKREAAKSHVI-YEKSGHLSYDNKIN 1227  
Db AS-POERSESDSPDKAKTRTPLRQRSGSGSPVDSKSLRPRRSRG---SSPEVKD 638  
Qy 1228 -----ARECTRS-----PETAHEISLKRSYVESVEGNIKOGMSRSPVAPLEGL 1272  
Db KPRAAPRAQSGSDSPPEKAPAPALPRRSKSGSSKGRGSPGEGSSTESSPEHPKSR 698  
Qy 1273 ICRALPRGSHSLDKERTVLSGSIHQGT-----RATESFEDGLKYPKOLKR 1320  
Db TARGSRSSPEPKTKSTPPRRSRSSRSPPELTRKARLSRRSRSSASSPFRSRTPPR-HR 757  
Qy 1321 ESPPIRAFEKA-----ITKGKPYDGITTIKEMGRSIEHPRDILTQESRKTPEYVQS 1373

Db 758 RPSVSSPEPAEKSRSSRRSSASSPRTKTTSSRRGRSPSPKPR-GLQSRGRSRREKTRT 816  
Qy TRPIEGSISOGTPIKFDNNSGSAIKHNKSLITGPSKLSRGMPLEIVPENIKVVERG 1433  
Db TRRRDRSGSSQSTRRRQRSSRSRVRTRRR---GGGYHSRSPARO---ESSTSRR 869  
Qy KYEDVKAGETVRSRHTSVVS-----SGPSVLSRSTLHEAP-----KAOLSPGIY 1476  
Db RRGSRSTPPTSRKRSRSTSPAPWKRSRASRATHRRSRSTPLISRRSRSTSP-VS 928  
Qy DTSARTPVSYQNTMRSRGSPMARTSDVTTPNKNSTNHERKSTLTPTQRESIIPAKSPV 1536  
Db RRRSRSTSVTRRSRSTRASPVSRRSRSTPP---VTRRSRSTPTTTRRSRSTP-- 983  
Qy GVDPVWVSHSPDPHHRGSTAGEVWVSHLPTQDPAHPHRLDPAAYLFOQLSPTPG 1596  
Db ---PVTRR-----SRSTPVTRRSRSTSPITR 1011  
Qy YPSOYQLYAMENTROTILNDYITSOOMOVNLRPDVARGLSPREQPLGLPYPATRGIIDL 1656  
Db RRSRSTSPVTRRSRSTSPVTRRSRSTSP-VTRRSRSTP-----PAIR---RR 1061  
Qy NMPPTILVPHPGGTSTPPMDRITVPGTQITFPFPYNSASMSGPHTHLAAAASAEER 1716  
Db SRSRTPLLPRKRSRSP-----AIRRSRSTP-----RTARGKR 1098  
Qy EREREKERERERIAAASDLYLRPGSEOPRPGSHGVYRSPSVRTOETMLQORPSVFQ 1776  
Db SLTRSPAIRRRSAGSSS---DRSRATPPATRNHSGSRTPPVALNRRSMCFRPSNSP 1156  
Qy GTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLPASRYTAADALAALYDAAAAPQMDV 1836  
Db TPLD-----RCRSPGMLPLGSSR-----TPMSVLQAGGS--NMDG 1191  
Qy SKTESKHEAARLENLRSAVSEOOOLEQKLEVEKRSVQCLYTSSAPPSKGPQPHS 1896  
Db PGRIPDHORTSVPENHAQSRALA-----LTAISLGTARPPPSM 1231  
Qy SVTYSEAGKDKGPP-----KRYEELRTRGKTTITAAFDIVITRQIASDKDARER 1950  
Db SAAGLAARMSOVPAVPLMSLRTAPAAANLASRIPAASAAANLASARTPAIPTAVNLADS 1291  
Qy GSQSDSSSSLSHRYETPSDAIEVI SPASSPAQEQKLTQYQEVKANKAENDPTROY 2010  
Db RTAAAAAANLASPRTAVAPSAVNLADPTPTAPAVNLAGARTPAALAALSILTSGET--- 1348  
Qy EGPLHHYRPOQESPOOQLPPSSQAEGMGVPRTHRLITLADHICQIITODFARNQVSS 2070  
Db -----PPTA-----ANYPSS 1359  
Qy QTPOQPTSTFONSPSALVSTPVR-----TKTSNRYSPESQAQSV----- 2110  
Db RTPOAPASANLVGRPSAHATAPVNIAGSRATAALAPASLTSARMAPALSGANLSPRVPL 1419  
Qy HHQPGSRVSPENLVDRSGSRPKSPERSHVSE--PYEPISPPQVQVYVHERQDSLLL 2167  
Db SAYERVSGRTSPP--LLDRASRTPPSAPSQSRMPTSERAPSSSRMGQAP----- 1467  
Qy LSQGAEPAPORNDARSFGSISYLPSPFTK-----LENTSPWKSKEIIFRKLNS----- 2218  
Db SQSLLPPAQ--DQPRSP-----VPSAFSDQSRCLIAQTTPVAGS-----QSLSSGAVA 1513  
Qy ---SGGD-----SDMAAOPGTEIFNLPAVTTSGS 2246  
Db TTSSAGDHNGMLVPAPGVPHSDVGEPPASTGAQOQPSAALQAPAKERRSSSSSSSSSS 1573  
Qy VSSRGHGFADPASNGLGLEDIIRKALMGSFQDKVEDHGVMSQPMGVVPGTANTSVTSGE 2306  
Db SSSSSSSSSSSSSSG-----SSSDSE-----GSSLVQPEYALKRVPSTP 1616  
Qy TRRE---EG---DPSHHGGVCKPKLISKNSRKSIPQGGYLGTERPSSVSVSHSEG 2360  
Db APKEAVREGRRPPEPT-----AKRRRRSSSSSSSSSSSSSSSSSSSSSSSSSSSS-- 1669



QY 2361 DYHROTGCWAWEDRPSSTGCTGFYFNPLTMRMLSTPTPIACAPSVAVNAQAHPQQNRW 2420  
 Db 1670 -----SSSSSS-----SSSSSPSAKPGPQALPASPAPKPPPG 1703  
 QY 2421 EREPAPL 2427  
 Db 1704 ERSLLPV 1710  
 RESULT 27  
 A47171  
 chondroitin sulfate proteoglycan PG-M core protein - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
 C:Accession: A47171  
 R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.  
 J. Biol. Chem. 268, 14461-14469, 1993  
 A:Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed during  
 A:Reference number: A47171; MUID:93300846  
 A:Accession: A47171  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-3562 <SHI>  
 A:Cross-references: GB:D13542; NID:g391643; PIDN:BA002742.1; PID:g391644  
 A:Experimental source: stage 22-23 developing limb buds  
 A:Note: sequence extracted from NCBI backbone (NCBIN:134456, NCBI:P134457)  
 C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin  
 F:166-243/Domain: link protein repeat homology <LNK1>  
 F:264-345/Domain: link protein repeat homology <LNK2>  
 F:3258-3289/Domain: EGF homology <EGF1>  
 F:3296-3327/Domain: EGF homology <EGF>  
 F:3334-3454/Domain: C-type lectin homology <LCH>  
 F:3461-3517/Domain: complement factor H repeat homology <FHD>  
 Query Match 3.04; Score 374.5; DB 2; Length 3562;  
 Best Local Similarity 18.14; Pred. No. 9.5e-07;  
 Matches 511; Conservative 379; Mismatches 1102; Indels 833; Gaps 118;  
 QY 9 NOGAFSTEO-SRYPPHSVQ-----YFPNTHQO--EFAVPDYRSSHLEVSQASQL 56  
 Db 781 SEGRFGSEKFTHTPGVMMLQDKQVYTEFTSHTKRIELDTEDDISGMEPTSSPGQI 840  
 QY 57 LQ-----QQQOQLRRRLSILSEFHFGSDRQERRTSYEPFHGPGS----- 97  
 Db 841 IEYTKHLGAPVASVTDKTSMTAETSEDEEVVSADFQTKGT-EVFTSSSLDEKF 899  
 QY 98 -----PVDHDS-----LESKRRLPEQVSDSHFQVSAALPLVHLPLEGLRA 139  
 Db 900 TLSKIPDESSATVKFSFSSSGTGLPVATVLEVTDEADETSGYVNLNFTSTPEGEQR 959  
 QY 140 SADAKDPAPGGKHEAPSSPIGQPCGDQNASPSKLSKEELIQSDMRDVRDEIAKVQOI 199  
 Db 960 KA-TEKSPA-----TSAEDEVSTGEISKYTMTEG-QQISSVTSAAKESV 1002  
 QY 200 LKAKKKQOQ-----LEEAAKPEPEKPVSPPEVE-----QKH 232  
 Db 1003 AALQEREQSVGLPTEKPEKFTDVTETITVPQREGDTSLPVTVVGSSEIDIGEMQVTDH 1062  
 QY 233 RSVIYIYDENRKAEEAHKIFEGLPKVELPLYNQPSDKVYHEN---IKTNQVMRKKL 289  
 Db 1063 TSFDSIIITEATVTSRASEV---PK-ELSTKQDDRELGTAMGSTLPVTSVMHEQKT 1117  
 QY 290 ILFPKRNRHAKREQKICORYDOLMEAWKKVDRI---ENNPRRKAKESKTREYEEKQF 346  
 Db 1118 TAGFESPQTITQEKHDEMGSAYDEMYPATLSVPALMLTEYGQVSGPVTSTSLHGT 1177  
 QY 347 PEIRKQEQEQERFQVQRGAGLSATARSEHEISEIIDGLSEQENNEKQMRQLSVIPPM 406  
 Db 1178 PKAETATDQEEKITEAVPVFTGQAKVYESK-----GTTTREDRDVGSNSVLPFH 1229



Db 2072 ---KDLTAETVSSPESV-----VNNSTLDMVTHGTIRAVAES 2110  
Qy 1414 SRGPPLEIYP-ENIKVVERGKVEDKAGETVRSRHSVSVSSGPS--VLRS----- 1461  
Db 2111 KKGKGSFSAVSLGKILMIEHGSGEELK-----VDSSTKLMSNGTEKLLGSHFFDOGS 2166  
Qy 1462 ---TLHEA- PKAQLSPGIYDTSARPTVSY-QNTMSRSGPMNRTSDVITPNK--S 1512  
Db 2167 GEATLTESATKASVP-----TKPPEQOYGRKTVSMPSAVVHA---YTAEPNELVTS 2218  
Qy 1513 TNHERKS--TLTPQRESIPAKSPVPGVDVPSVSHSPF--DPHHRGSTAGEVYWSHLPTQL 1568  
Db 2219 TEHDITSLQVTDTEMEKAANELT--VTSFATNPLPSEDVHS-----WEDRPREI 2267  
Qy 1569 DPAMPFHRALDPAAAA---YLFQRLS-----PTCPYPSOYQIYAMENTROTIL- 1614  
Db 2268 LP-----KATESGEATEDFFISTQANHEHVEFLSVPTTRPHEENKVAESDEKILLP 2322  
Qy 1615 --NDYIT-----SQOMOVNLRP--DVAR-GLSPREQPLGLPYPATR 1650  
Db 2323 FNDRVTVESAVERKYLSSPFTDEQEELVQNIFPTEDIPRLFTPKEE----- 2372  
Qy 1651 GIIDLTMNPT--ILLVPHPGTSTPPMDRTYIPG-----TQITFPPRPYNSASMS 1700  
Db 2373 -----KPTNNELISDPLFSGQSGDEFTVIPSVESLAVKETNTLSPPWHPFASVGP 2424  
Qy 1701 ---CHPHTLAAAASAEERE--REREKERERERIAAASDLYLRPGSQPGRPGSHGVRS 1756  
Db 2425 KLSTDKQVTEGSTGDSNAEINEITTAELTAYSMATSSPALBEESSSHSNDKDD 2484  
Qy 1757 PPSVRTQETMLQORSPVFGTNGTSVITLDPDPTAQLRIMPLPAGGPSISQGLPASRYNT 1816  
Db 2485 ITHYFLVIEDPYNKEMDHRRGNGTS-----RPLTPG--DVSLEESSHML 2528  
Qy 1817 AADALAALVDAASAPQMDVSK-----TKESKHEAARLEENLRSGA--AVSEQ-----Q 1864  
Db 2529 TTDDVTPVSVILSETPYLEMGKSLATSATKMPSRVLPSESGEGWGVSDSFAPDTLTH 2588  
Qy 1865 QLEQKTLVEKRSVQCLYTSAPFSGRPQPHSSVVYSEACKDKGPPPKSRVYEEELTRGK 1924  
Db 2589 STAPSVMEVE-----ITASHIPG-----VYSEVMTHTVPGDGS-----Q 2623  
Qy 1925 TTIT--AANFID--VIITROTASDKDARERGSSDSSSSLS----- 1962  
Db 2624 TVITGLASLFTKEKEIVANRTAADPKGTSEELTSDTGMSLDILPVVDRRHVTLNVS 2683  
Qy 1963 -----SHRYETPSDALEVISPASSAPPOB--KLQTYQPEVVKANAENDPTR----QY 2010  
Db 2684 GDITLIERLQIPSEKTTIIDMDHRSKMPEDILSVQTMPLNIVIRSTQVSDDNMKAEDKY 2743  
Qy 2011 EGPLHYRPOES-----PSPQOOLPPSSQACMGQVPRTHRLITLADHICOLIITDFA 2064  
Db 2744 DSIINFSTVENSPGSGDNLSTLTSIQPSSESVTAGHGKLVKDLGSGVAMQAFATLT 2803  
Qy 2065 R-----NOVSSQTPQOOPPTSTFQNSPALSALVSTPVRTKSNRYSPESQASV 2110  
Db 2804 TTVLNLGILFPTVPSLVSPHMPHESKESEFEAKHIGTST-----TDDVTEPYTSANN- 2857  
Qy 2111 HQRPGRSVENLVKSR-----GSRPGKSPERSHVSEPEYPTSPQVPPVHKKQ----- 2162  
Db 2858 -----QVITDQSKTMSISGFGMGQEEG--DKKPMIPSLTDLTMEKALT 2905  
Qy 2163 --DSLALLSORGAPAE-QRNDARSPGISVL--PSFTKLENTSPMVKSKQEIFRKLN 2217  
Db 2906 TFDVSMVTTQSMQSHAVTSSSSSEKSTHYVMQTKASTEYETDSV-----SLN 2955  
Qy 2218 SSGGSDMAAAOPGTIFNLPAVTTSGSVSSRSHSPADPASNLGLEDIIRKALMGSD 2277  
Db 2956 SVSONPKASSVTVLVNGVSKYKPEVILPSTSAKSDSDSDHSD-----GTKE 3003  
Qy 2278 KVEDHGVVMSQPMGVPGTANTSVVTSGETRRE--EGDPSPHSGGVCKPKLISNSRKS 2335

Db 3004 VSDMAATYKPPPTDLDLTTVSSLLVSPESPESISTESTPH-----FNKFVTERS 3054  
Qy 2336 KSPIPGQGYLGTERPSSVS-----SVHSEGDYHROTPGWAMEDRPS-----STGSTQPP- 2384  
Db 3055 EETESSVNDLIIENAVTSGDSPSIH---DY--PTAFWNGERTSTDPVKLSTIEVEFSS 3109  
Qy 2385 ---YNP-----LPMRLMSSSTPPTPIACAPSAVNOAA--PHQONRIWEREPAPLLSAQYE 2433  
Db 3110 ERVKNPQESDRSTERPRLLSAPVSDSPNSIEGVFKPDQ-----EAVTMLTSSLE 3162  
Qy 2434 TLSDS 2438  
Db 3163 PLDRS 3167  
RESULT 28  
T20531  
hypothetical protein F07A11.6a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20531; T2776  
R:Palmer, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: Z19287  
A:Accession: T20531  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2526 <WIL>  
A:Cross-references: EMBL:Z66511; PIDN:CAB54210.1; GSPDB:GN00020; CESP:F07A11.6a  
A:Experimental source: clone F07A11  
R:Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20417  
A:Accession: T2776  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2526 <W12>  
A:Cross-references: EMBL:Z69904; PIDN:CAB54501.1; GSPDB:GN00020; CESP:F07A11.6a  
A:Experimental source: clone ZK20  
C:Genetics:  
A:Gene: CESP:F07A11.6a  
A:Map position: 2  
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 1286/3  
Query Match 2.9% Score 372; DB 2; Length 2526;  
Best Local Similarity 18.2%; Pred. No. 7.7e-07;  
Matches 467; Conservative 343; Mismatches 898; Indels 860; Gaps 115;  
Qy 200 LKLLKKQOQLEEE-----AAKPPPEKPVSPPPVEQKHRSIVQIYDENRKAEE----- 249  
Db 64 LRARSTRQPEDSTEXKISAVIPEPTQN-SPIP-EFPSESTACVVYEIQSGSTPERDLFE 121  
Qy 250 ---AHKIFEGLGKVELPLYNQ--SDTKVYHENIKTNQVMRKKLILFF---KRNHARK 301  
Db 122 LVKHKSRSGVPIDILESTTEPGMKKARVHYRLDTGLKADKSLILGRPPKFRVYPT 181  
Qy 302 QREOKICORYDQLMEAEKKVDRIENPRRKAKESKREYVEKOPPEIRKQREOQERFOR 361  
Db 182 SGEOK----HPQCHPSYSYAIPLKLGDHLLKASCSVHVPHLDHRSPD-----HYRRRRES 232  
Qy 362 VQORGAGLSATIAKSEHSEIIDLSEQENNEKQMRQLSVIPPMFDAQOR---VKFI 418  
Db 233 YGQ---VIDVDMKSNDKAPAVVQFTNIDDAQALQDNTNIPKPMYSQSRSHRIIFYL 289  
Qy 419 NMGLMEDPMKVKY-----DROFMNVWTDHEKE--IFKDKFTQHPKNFGLI 462  
Db 290 PIECTNEEIMLIIRSLSDRIVDICVDWMDRSAVITLDDMEPANLLLRKMLVGRNFG-- 347  
Qy 463 ASYLERKSVDP-C---VLYYLTTKKNYKALVRNRYGKRGRNQOQIARPSQEKVEKE 518  
Db 348 ---EHRVAVDFCSDRPNLFIYFNKKNENIEVAARSSPTSSENDQGSPPSSSRDRONL 403







Db 2247 DSMGEGSAFRILRSSTMGNSGSPSASGT-----TSPSTSSSIS 2289

## RESULT 29

RBHUAP

adenomatous polyposis coli protein - human

N;Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1991 #sequence.revision 31-Dec-1991 #text\_change 21-Jul-2000

C;Accession: A37261; B39658; A44928; A49319; I54271

R;Kinzler, K.W.; Nilbert, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith

chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.

Science 253, 661-665, 1991

A;Title: Identification of FAP locus genes from chromosome 5q21.

A;Reference number: A37261; MUID:91335210

A;Accession: A37261

A;Molecule type: mRNA

A;Residues: 1-2843 &lt;KIN&gt;

A;Cross-references: GB:M74088; NID:g182396; PIDN:AAA03586.1; PID:g182397

R;Joslyn, G.; Carlson, M.; Thilveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grod

arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Le

Cell 66, 601-613, 1991

A;Title: Identification of deletion mutations and three new genes at the familial polypos

A;Reference number: A39658; MUID:91330307

A;Accession: B39658

A;Molecule type: DNA

A;Residues: 1-183, 'L', 185-969, 'N', 971-1308, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P'

A;Cross-references: GB:M73548; NID:g190163; PIDN:AAA60354.1; PID:g190164

R;Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst

Cancer Res. 52, 643-645, 1992

A;Title: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a c

A;Reference number: A44928; MUID:92119623

A;Accession: A44928

A;Molecule type: DNA

A;Residues: 1506-1525 &lt;MIK&gt;

A;Cross-references: GB:S78214; NID:g243541; PIDN:AB21145.1; PID:g243542

A;Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBI:P:78218)

R;Spurio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber

Cell 75, 951-957, 1993

A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.

A;Reference number: A49319; MUID:94073973

A;Accession: A49319

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 'G', 143-171, 'P', 173-179 &lt;SPI&gt;

A;Cross-references: GB:S67787; NID:g461061; PIDN:AAD13997.1; PID:g4261697

R;Lambertz, S.; Ballhausen, W.G.

Hum. Genet. 90, 650-652, 1993

A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly

A;Reference number: I54271; MUID:93186137

A;Accession: I54271

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-4 &lt;LAN&gt;

A;Cross-references: GB:S56365; NID:g266243; PIDN:AAD14918.1; PID:g4262770

C;Genetics:

A;Gene: GDB:APC

A;Cross-references: GDB:119682; OMIM:175100

A;Map position: 5q21-5q22

A;Note: mutations of this gene can result in familial adenomatous polyposis or sporadic

C;Superfamily: adenomatous polyposis coli protein

F;1-730/Domain: leucine-rich &lt;NTD&gt;

F;7-727/Region: coil #2/status predicted

F;185-227/Region: coil #status predicted

F;731-2832/Domain: serine-rich &lt;CTD&gt;

F;1131-1156/Region: acidic

F;1558-1577/Region: acidic

F;1866-1893/Region: highly charged

Query Match

Best Local Similarity 2.9%; Score 365; DB 1; Length 2843;

19.0%; Pred. No. 1.7e-06;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKQEQE-RFORVGORGAGLSATIRSEH-----EISEIIDGLSEQENNEK 395  
Db 981 PSIESYSEDESCKFCGYPADLAHKIHSANHMDNDGELDTPIYSLKYSDEQLNSGR 1040  
QY 396 QMRQLSVIPPMFADAEORRVKFINNHGLMEDPMKVYKDRQPMYVTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPQNERWARPKHIEDEIKQSEQRQSRQSTTVPVYTESTDDKHL 1087  
QY 454 QHPKNFGLIASYLERKSVDPDCLVYYITKKKNYKALVRRNYKRRGRNQOIAJPQSEK 513  
Db 1088 KFOPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQNVSLQCE- 1132  
QY 514 VEEKEEDKAETKK--EEKKDEEKKEDSKENTKDKIDGTABETEEREQATPRG 571  
Db 1133 -DYEDDDKPTNYSERYSEEEQHEEERTNYSIKYN-EERHVDOPIDYSUKYATDIPSS 1190  
QY 572 R-----KTANSQGRKRGRITRSMTNEAAAS-----AAAAATEEPPLPPP 614  
Db 1191 QKQSFSEKSSGSSGSSKTEHSSSENTSTPSSNAKRONLHPSSAQSRSGQPKAATCK 1250  
QY 615 PEPITSTPEVET-----SRTEEMEVAKKGLVEHGRN-----WAAIAKM 653  
Db 1251 VSSINQETIQTICVEDTPICFSCSSLSLSSAEDEIGCNOTQEADSANTLQIAEIK 1310  
QY 654 VGTKSE-----AOCKNFYNYKRRHNLNLOOHK-----OQTSR 688  
Db 1311 IGTSAEDPVEVPVAVSOHPKTKSRIGGSSLSSESRHKAVERFSSGAKSPKSGAOTPK 1370  
QY 689 KPREE-----RDVSOQESVASTVSAQDEDEIASNEEENPEDESEAVKPSE--DSPEN 740  
Db 1371 SPPEHYVQETPLMFESRCTSV-SLSDSFESRSIASSVQSEPCSGMVSGIISPDLPSPGQ 1429  
QY 741 A--TSRQNTPEAVELEPTTETAPSTPSLAVSTPKPAEDESVEVQVNDISAEATQMDV 798  
Db 1430 TMPFSRKTTPP---PPQATQKREVKNKAPTAEKRESGPKQAANAQV--RVQVLPDA 1485  
QY 799 DQOQHSABE-----GSCV-----DPPATKADSVDEVR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHFATSTPDGFCSSSSLSALSILDEPFIQK----DVELRIMPVQENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDEDLVVAQOINAQRPQSDNDNSATCSADEQVD----- 890  
Db 1542 PKESNQEKEAEKTIQSEKDL-----DDSD-----DDOIELEECIIS 1581  
QY 891 --GEPEQRQMFMDSPSLNPTGSLVSSPLKPNPLD---LP-----QLQHRRAVIP 938  
Db 1582 AMPTKSSRKAKKPAQATASKLPPPVAR-----KPSQLPVYKLLPSQNRLQPKHVSFT 1634  
QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRYVCEGTPINSTATSLDITIEPPNELAAGEVGRGAQSGEKEKRTITFEG 1694  
QY 971 AL-----LEEQRQREQIDLECRSSSTPCGTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGKTSSTVTIPELDDNKAEGDILAECINSAMPKSKHFKVRVKIMDQVOAS 1754  
QY 1008 --QAPQLITNLPEGVLLPTTRTPRPPPLI-----PS 1039  
Db 1755 ASSAPNK---NOLDGKKKKKTPSPVKPIONTEYTRVRKNADSKNNLNAERFVSDNKS 1811  
QY 1040 SKTTVASEKPSF-----IMGSGTS-----QGTPTGYLTSHNOASYTQETPK 1080  
Db 1812 KQONLKNNSKDFNDKLPNNEDRVGRSFAFDSPPHYTIEGTP--YCFSRND----- 1860  
QY 1081 PSVGSISLGLPROQESAKSATPYIKQEFSPRSQNSQPEGLLVYRAQHEGVYRGTAQAI 1140  
Db 1861 -SLSSLDFODDVDLSREKAEELKAKENKESAKVTSHTE--LTSNQOSA---NKTQAI 1914  
QY 1141 EGSITRGTPTSKTSVESIPSLRGSITQGTALPQTGTPTEALVK-----GS 1186  
Db 1915 QPINRGOPKPILOKQS-----TFPOSSKIDPRGAATDEKLQNFATIENTPVCFSHNS 1968







QY	2001	QAEKD---PTROYEGPLHHVYRQOESPSPQOQLP----	PSSQAEGMG-----	QVPRTH-	204				
DB	1615	QVQRIIPIEVEQAPT---	IPQPPRPAPKSELPKAKVAPLDDSKSRVREAPLNILKGRITS	1671					
QY	2047	-----RLITLADHICILITQDFARNQVSSVOTPOOQPTSTF-----	QNSPSALV---SIPVR	2094					
DB	1672	BEQQKELVESLERPLTIITQ-----	OKPPEKPTEDIGALSPLSPTNTLAEVEEYPM	1722					
QY	2095	TKTSNRYSPESQAQSVHH-----	QRPGRSVSPENLVDSKRSRPGKSPERSHVSSEVP	2149					
DB	1723	DMQSVPHSPQEKQEEIEALSEIIIEEPQANKVEKPV-----	SAEKDNESLEAPEI	1774					
QY	2150	ISPPQVPVVIHEKQ-----	DSLILLSQGA-----	EPABQR	2179				
DB	1775	INEPIRVLVETKIMGPGKSLNEDNDDDDGGSECLDSIGDLISRTIQRFNTSIDDPSSIR	1834						
QY	2180	NDARSPGISYLSFETKLENTSPMKVKKQEIFRKLANSGGD-----	SDMAAAQPCTE	2234					
DB	1835	DSFSSISFGDROKFRTAENIRHODLLPQSSVQYLKSSPNPQSOQLLVTLNLSMDSPSDL	1894						
QY	2235	IFNLPAY-----	TTSGSVSSRG-----	HSFADPASNGLGEDIIRKAL	2271				
DB	1895	SPNAPPVGFENTAQFLKQLQOEDRPSAEGSIDSSGFVKVDEHGLDEFAAPPVHPDMQKSV	1954						
QY	2272	MGSF--DDK---VEDHGVVMSQMGVVPQTANTSVVTSGETFRRE-----	GDSP	2316					
DB	1955	FGSLGSDMKPGSQDDGFVF-----	IERNEANEATLKKNQKMSHHNDVIEKNTFYENDNAP	2009					
QY	2317	HSGGVCKP-----	KLIS---KNSRKSKSPIPGOGYLGE-----	RPSVSS	2355				
DB	2010	TAALLESPIAEARKLVQDAVESASEYKQAVDSGDEIGRELLDNVQKIEQVKEPIDS	2069						
QY	2356	VHSE-----GDYHRTQPCWAMEDRPSSTGSTQTPYNPPLTMRMLSSTPPPPIACAPSAVNOA	2411						
DB	2070	LHKAYDGVGFVHETVPNAVDDPVRE-AEKQLPESPVPKEI---ETPEPLVDIHDVTDKV	2125						
QY	2412	APHQONRIWEREPAPLLSAQYET-----	LSD	2437					
DB	2126	HDEVDFNL-RREPTP-----	PFETDDVAPLSD	2151					
RESULT	31								
RESULT5	hypothetical protein W06A7.3a - Caenorhabditis elegans								
C:	Species: Caenorhabditis elegans								
C:	Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999								
C:	Accession: T26215								
R:	Ainscough, R.								
	submitted to the EMBL Data Library, August 1996								
A:	Reference number: Z20173								
A:	Accession: T26215								
A:	Status: preliminary; translated from GB/EMBL/DBJ								
A:	Molecule type: DNA								
A:	Residues: 1-2607 <WIL>								
A:	Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN00023; CESP:W06A7								
A:	Experimental source: clone W06A7								
C:	Genetics:								
A:	Gene: CESP:W06A7.3a								
A:	Map position: 5								
A:	Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2;								
Query Match									
		2.9%	Score 361;	DB 2:	Length 2607;				
		Best Local Similarity	18.1%	Pred. No. 2.1e-06;					
		Matches 428;	Conservative 320;	Mismatches 937;	Indels 686; Gaps				
QY	509	SOEKEVEKEEDKAEKTEKKE-----	BEKDEBEKEDEKEDSKENTKEKDIDGTA	558					
DB	25	SDQESSENKEPDASENVEETPVSSAKVIALESSGDCCEGNKIKISANENSVPPDGAKPA	84						
QY	559	-----	EFTEEREQATPR-----	GRKTANSQGRRKGR	585				
DB	85	HTEQIEPTIILPSETVTOLEDHLSSEKNEVESSIKSEKNVIGLSDGDTFRDPTSG	144						



Qy	586	TRSTWNEAAA-----	ASAAAAATEE-----	pppp-----	610
		:     :	:	:	
Db	145	APABEDNAEATITDTFVPLESAGDIPSENEIKEVASADPVGLEEYIIGNIPNAPVND			204
Qy	611	LPP--PPEP1STEPVETSRKTEEMEV-----			635
		:     :     :     :			
Db	205	IPNVFTPEVANDETVEFVSVAEASIPVVVELEPIGDEYFQRPVENFSPDNINLEE			264
Qy	636	--AKKGLVEHGRNWAALAKMVGTKSBAOCKNFYFKRRH-----	NLDNLQOQHKKQTSR		688
		:     :     :     :	:     :	:	
Db	265	SGAEQVULEN-----	NMFTPLDELPQOQKILNEKAHEIEASGDEFIKDHPFPVEN		316
Qy	689	KPREERDVSOCEASVASTVSAQED-----	EDIEASNEENPEDESEYAEVKPSEDSPEN		740
		:     :     :     :	:     :     :	:	
Db	317	EQNVNEQSSRVETVHSFIGLESSGVGLTGAVSDSVANNVKENTESPDIISLEASGDELSK			376
Qy	741	ATSGNTEPAVELEPPTETAPSTSPSLVPSYK-----	PAEDSEVETQVNDISISA		790
		:     :     :     :	:     :	:	
Db	377	LVE-----AREIITESKDAYSTD-----	VPESRKTVSDVIGLEEAGDKIVSNVSNVGN		436
Qy	791	ETABQMDVDOEHSABEGSVCDPPPKATKADSVDVVRVPENHASKVEGDKTKERDLDRAS			850
		:     :     :     :     :     :	:     :     :	:	
Db	427	PDESQLALEGQDNVPE-----	LPIENSEQETVAVKATSCDHVDSQKTLERASSLEE		479
Qy	851	EKVPRDEDILVAAQINAOARPEPQSDNDS-----	SATCSADEDVDGGEPR		895
		:     :     :     :	:     :	:	
Db	480	DIMSPENVLGITSSQTLSDYLPVISEDODSIPVPVTEVEETSEKLVKASSLEEDV-VSPE-			537
Qy	896	QRMPPMDSKPSLLNPFGSILV--SSPLKNPLDLPOLQHRAAVIPPMVSVCTPCNPIGT			953
		:     :     :     :	:     :     :	:	
Db	538	--VLELDDRYQNKPNSEVETAVDASKTEGDFSQSDPSRATETFEKLVTVTENLLPAGDK			595
Qy	954	VSGVALYQRIHKAMHESALLEEQRQEQIDLECRSSSPCGTSKSPNREHVEVLQAPHQ			1013
		:     :     :     :     :	:     :     :	:	
Db	596	LS-----EERIQIEIRENETISQ--PGKEEDDLE-----	NANDPDDE-TIVEKIVSM		638
Qy	1014	LITNLPBGVRLPNT--RPTRRPPPLIPSKTTVASEKPSFIMGSGISQGPYGLTSHN			1070
		:     :     :     :     :	:     :     :	:	
Db	639	AESSLP1EAVISTEDGTSQDPAQNAIPDSEET-----	TVDDSQTEEIFTDDN		686
Qy	1071	QASYTQBTTPKPSVGSISLGLPQOESAKSATLPYIKQBEFSPRSONSOEGLLVRAHQEG			1130
		:     :     :     :     :	:     :     :	:	
Db	687	VKKSKEPTPAENDTINYLPGGEGPEDNAEK--RNEAVSPNDTSE-----	IKQDLEN		739
Qy	1131	VVRGTAGAIQGSITRGTPPTSISVIESIPLSRGSTQGTALPQT-----	GIPT		1179
		:     :     :     :     :	:     :     :	:	
Db	740	LENGASGPDNNVQVDBAAQEDPTDPTVDVETTSKSENPKAPTEDDNATEFTEAGLET			799
Qy	1180	-----EALVKGSIRMP1ED--SSPEKGREEAASKGHVIEGKSGHIL			1220
		:     :     :     :     :	:     :     :	:	
Db	800	TEAFGDAEHVSYLDANIEKLIVAMADEPLPVDELVSIEERPEVAPAES-----	TGEDEDIF		855
Qy	1221	SYDNIKNAREGTRSPTRAHE-ISLKRSVYESVGNIKQGMRESFVSPASPLEGLICRALPR			1279
		:     :     :     :     :	:     :     :	:	
Db	856	RDRRTVSLTGTGDQNAPIQVIFVGDGDNPDANADQ-----	ER		894
Qy	1280	GSPHSLKERTVLGSGIMOGTTPRATESPEDGLKYPKOIKRESP-----	PIRA		1327
		:     :     :     :     :     :	:     :     :	:	
Db	895	TSEHNELIESD-----KESEBAITKNEEDVDQDP--IQSEEP1T	SOEGESSIGNKIVA		945
Qy	1328	FEGAI1TKGK--PYDGITT-----	IKEMGRS1HEIPROD-----	ILT	1361
		:     :     :     :     :	:     :     :	:	
Db	946	VVGSVLLGGVAIPYGVLANENEDAHADREVEETGDS1RDRPEETFVSK1TSMVENILP			1005
Qy	1362	QESRKTPEVVQSTRPIIEGSIQGTPIKPDNNNGQSA1KHN-VKSLITGPGSK1SRGMPL			1420
		:     :     :     :     :	:     :     :	:	
Db	1006	STNDENPEAVSMVENYLPVNTGLEDSKEDNPDAPTAEAHSGEKNLRNDK1TTRDRDPE			1065
Qy	1421	E1VPENI-KVYBERGYEDVKAGETVRSRHS1VSVSSGSPVLSR1LHHPAQ1-LSPCIYD			1477
		:     :     :     :     :	:     :     :	:	
Db	1066	E--ET1LNKLVE-----	NALPTGV1TGSFTFV-----	SAPDAQELDE1TVVD	1103

Qy	1478	DTSARRTPVSYQNTMGSGPMNRISDVITIPNPKTNRHKKSTLTPQRRSIPAKSVPG	1533
Db	1104	HAG-----QNDTSEVEDAPEKSGAGGTVEKFTSMTESTILPQAP----	1142
Qy	1538	VDPVVSHSPDPHRRGSTAGEYVWSHLPTQLDPAMPFHRLDPAAXAYLFLQRLSPPTGY	1597
Db	1143	TOPENAESHIQDTGG--ASEI-----KDONQP-----EFSACHQGFVEYSAEPDQ	1188
Qy	1598	PS-QYQLYA-MENTROTILNDYITSOOMOVNLRPDVANGLSRPROGLPLYPATRGITDL	1655
Db	1189	ESAEFQLEAKDQDKETIENSEDAKETVWEKLVSLVENILPFAVLPSDVTYTKNSEDK	1248
Qy	1656	TNMPPTILVPHPGGTSTPPMDRITYIPGQITFPRPY-----NSASMSGH	1702
Db	1249	KELETQELSSKEITSCQP-----EYVPETSEAFVSDPEIFQVRKRASSTEBTKQTEPHA	1304
Qy	1703	PTHLAAAASARERE-----REREKERERTAAASDLYLRPGSEQPRGPGSHGY-----	1753
Db	1305	PIFTVGOSTEDDEQCSANWIDELVHDEDDKPYEVYANISVASENIDDDSTANAVPKTE	1364
Qy	1754	VRSPSPVRTOEMLQORQSVFQGTNGTSTVITPLDPTAQLRIMPLPAGGSIISQGLPASR	1813
Db	1365	VSSQLQVATVEFELGSAPE-----EESAIPAPEVQELSEKVEVQ-----PDLSONSPAPH	1414
Qy	1814	YNTAADALALVDAAASAPQMD-----VSKTKESKEHAAARLEEN-----	1852
Db	1415	-----KIIDLHFNIPKDHYDGYVPPGTSESSESKADQNGENQEEEDVYVAELN	1465
Qy	1853	-----LKSRSAAVSE-----QOOLEQKTELV-----EKRSVQC	1880
Db	1466	FHPTRQWRDEVTLSQASLSLVAEVCITDVDASDVNQDEESTLKLKLVVSPSPSLLEL	1525
Qy	1881	LYTSAPSPCKOPHSSVVYSEAGDKGPPPKRSYEEELRTGKTTITAAFNDFVITRQ	1940
Db	1526	DFTNDPKVIHPIPLME-----PATMTLEEM-----VEMIIADAVKESEME	1567
Qy	1941	IASDKARERGSSDSSSSLSHRYETPSDAIEVISPASSPAPPEQKLOTQOPEVVKAN	2000
Db	1568	VVTEISEMAPQVSESTCPIP-----EPLADLKLPEVDEK--TPEPEPVPG	1614
Qy	2001	QAEND--PTRQYEGPLHHYRPOQSPSPQOQLP-----PSSQAEGMG-----QVPRTH-	2046
Db	1615	QVORRIITPIEVOAPT---IPORPPRAPKSELPYAKVAKPLDDSKRSRFAPLNKLGRYTS	1671
Qy	2047	-----RLITLADHICOIITODFARNQVSSOTPOQPTSTF-----QNSPSALV---STPVR	2094
Db	1672	EEQKELVESLERPLITIQ-----OKPEKPTEDIGALSPLENTLAEYEEVPM	1722
Qy	2095	TKTSNRYSPESQAQSVHH-----QRPGSRVSPENILVDKSRGPKGSPERSHVSSEYP	2149
Db	1723	DMQSVPHSPQEKQEEIALSEIIEEPQAMKEVKPE-----SAPERKNESLEAPEI	1774
Qy	2150	ISPPQVPVHIEKO-----DSLILLLSQRGA-----EPABQR	2179
Db	1775	INEPIRRVLVETKIMGPGKSLNEDDDDGSECLDSTIGDLSERTIQRFNTSIDDPSSIRR	1834
Qy	2180	NDARSPGSIYLSPTFKLENTSPWVKSQKQIEPRKLSNCGGD-----SDMAAAAPGTE	2234
Db	1835	DSFSSISSFGDRQKFTATENIRQDLLPQSSVSYQLKSSPNPQQLLVNLNMDSPSDL	1894
Qy	2235	IFNLPAV-----TTSGSVSSRG-----HSFADPASNGLGLEDIIRKAL	2271
Db	1895	SPNAPPVGFENTAQFLKLEQEDRPSAEGSIDSSGFEKVDEHGLDEFAAPVPHDPQKSV	1954
Qy	2272	MGSF--DDK---VEDHGVVYMSQPMGVVPGCTANTSVVYTSGETRREE-----GDSP	2316
Db	1955	FGSLGSDMMKPGSQDDGFV-----IERNEANEATLKKNQKSSHHNDVIEKNYFNDNAP	2009
Qy	2317	HSGGVCKP-----KLIS---KSNRSKSPIPGOGYLGTE-----RPSVSS	2355
Db	2010	TAALLSPITAEARKLVLQDAVSEASYYKQAVDSGDEIGRELLDNVQKLEQVKPELVDS	2069
Qy	2356	VHSE---GDYHROTPGAWERDRPSSGTSTGTFPPNPLTMRLMSTPTPTACAPSAVNOA	2411











Db 863 SMWITQSPVSPQSTSPVSSSTTPSGPGLTLTSTSPSPSQSTTIGTQSTSPGISTTS 922  
Qy 1263 SPVASLEGICRALPRGSPHSDLKERTVLSGIMQGTTPRATTFESFEDGLKPKQIKRES 1322  
Db 923 EEMTS -QG---STQTPGSGTSTVQTPVSDSTSGS----- 955  
Qy 1323 PPIAFEGAIIKGYDGIITIKEMGRSHEIIPRODIITQESRKTPVQVSTRPIEGSI 1382  
Db 956 -----TWTVGSPGESSPIPTSQNTN--PSTSSGSSMSTQTPQSSQSTP-VESST 1004  
Qy 1383 SOGTPIKFDNNSGQAIKHNKSLITGSKLSRGMPPLIPEVNIKVVGRKYEDEV-RAG 1441  
Db 1005 SGAT-----SSSGSGT--TLTISPSPSSTIGSSOGSTSPVVSTISQGSTETPGTG 1057  
Qy 1442 ETVRSRHTSVSSGSPVLSRSTLHEAPKQALPGIYDDTSARRTPVSYONTMGRGSPMNR 1501  
Db 1058 STV-TKPTVSGSAGSGTATMGTEASTSG-----SSTSPNPSQST---SP---S 1103  
Qy 1502 TSDVTIPNKNSTNHERKSTLTPTQRESTPAKSPVPGVDPVVSHPDPPHHRGSTAGEYVW 1561  
Db 1104 TSGATSSPGSGT--TLTISPSPSSTIGSSOGSTSPVVSTTSGDMTSGST----- 1155  
Qy 1562 SHLPQLDPAMPFHRALDPAAAYLFQRLSPTPCYPQYQYLIYANTRQIILNDIYTSQ 1621  
Db 1156 -QIPGSGTGV-----TOPSTGSGSTSGEITSOGSTQTPRSSLSSTP 1198  
Qy 1622 OMQVNLRPDVARGLSPREOPLGLYPAT-RGIIDLTNMPPITLVPHPGCTSTPPMDRITY 1680  
Db 1199 AISTSTQSVSTN-SFGST---VTQPSIVRG---STSGSVTVTSGSTSGSSSATS 1251  
Qy 1681 IPTGTOITPPRPYNASMPGHPHTLAAASAEERERERERERERERERERERERERER 1740  
Db 1252 LSSSS-----PVPSTQSPNPST-----SGSSTPTNP 1279  
Qy 1741 G-----SEQPGRGSGHYVRSPF-SVRTOETMLQORPSVFOGTNGTSVITPLD--- 1788  
Db 1280 SOSTSPVSTTGTGEMTSHGTSQTPTIGSTVTO-----PSTVSGSSSGSTVITGSGE 1332  
Qy 1789 -----PTAQLRLMPAGGSPISQGLPASRY-----NTAAD-----ALAAVDAAS 1830  
Db 1333 ASTGSGSFKTSPSSISPVPTSP-----IPSTTFASSTSGSTISDVSSVSTSLAPLSS 1387  
Qy 1831 APQ-----MDVSKTKESKHAARLEENLRSAVSEQQOLEQKTLKVRKSVQCLYTS 1884  
Db 1388 LPSTVPSSTQSFSTSGSSKASSPVPSTQSTPTNPTGSTEFTLLSSTISGTOHTT 1447  
Qy 1885 -SAPSGPQPHSSVYSEAKDKGPPKRYEELRTGKTTITAAFNIDVIITQIATAS 1943  
Db 1448 MSKASSGSTSPSTN---SQTG-----STVTMGSSSTSGVSTSSAS 1484  
Qy 1944 DKDARERGQSSDSSLSHSHRYETPPSDAIEVISPASSAPPOE--KLQTYQPEVVKANQ 2001  
Db 1485 STOPQNSTSQSSAGSTVAS-----STASPAASTAPSTGTMSTSSSTSGTVGSTI 1534  
Qy 2002 AENDPTROYE-----GPLHHRPOQSPSPQOQLPPSPQAEQMGQ----- 2041  
Db 1535 SESSVTASASQGTGTVTMGSSSTSGVSTSSASTQPMQSTSGSAGSTVASSTAGLVS 1594  
Qy 2042 ---VPRTHRLI-----TLADHICQIITQDFARNQ-----VSSOTPOQ 2076  
Db 1595 TSTVPSSTGTMGSSSTSGVSTISSTSTASASSQSTGTVTMGSSSTSGVSTSSASTQ 1654  
Qy 2077 PTSTFQ-----NSPSALVST---PVRTKTSNRYSPESQAOQSVHHQPGSRVSPENL 2124  
Db 1655 QMSTSGSSAGSTVASSTGLVSTSTVPSSTGTMGSSSTSGVSTISSTSTASASSQGT 1714  
Qy 2125 VDKSRGSRGKSPERSHVSSEPEYIPSPQVPPVHHEKQDSLILLLSORGAEPAEQRNDARS 2184  
Db 1715 STVTMGSSSTSGVSTSSASS-----GQPMQSTSGSSAGSTVVSST-ASPAASSTAPSS 1767  
Qy 2185 PGSTSYLPS-----FTFKLENTSPMVKSKQEIFRKLNSSGGSDMAAOPQTEINLPAV 2241  
Db 1768 TGTWSSSTSGVSTMSQSSAASTTSHTGTSTVTLGSSSTSSNQMSTSQ-GSSVSGSTVAS 1826

Qy 2242 TTSQSVSSRGHSFADPASNGLGLEDIIRKALMGSFDDKVEDHGVVMSQPMGVVP-GTANTS 2300  
Db 1827 STAGLVST-----STVPSSTGTMGSSSTGTSSTTASASSQGTSTVTMGSSSTS 1881  
Qy 2301 VVTSQETTRREGDPSPHGGVCKPKLISKSNRKSKSPIGO-GYLGTFRPSSSVSHVEE 2359  
Db 1882 GYSTSASSTQPMQSTSGSSAGSTVASSTAGLVSTSTVPSSTGTMGSSSTGTVGTSISE 1941  
Qy 2360 GDYHRQTPCWAWEDRPSSTGST-----OPFYNPLTMRMLSSPTPTPIACAPSAVNOAAHP 2414  
Db 1942 SSTAASST-----SSQTGSTVTIGSTGNPSPRSLSLQITTP-----SPS-----OSTES 1987  
Qy 2415 QONRIWEREPAP-----LLSAQYETLS 2436  
Db 1988 TOTSLPSSSPSTHVSSEGTTMS 2013

RESULT 34  
B35098  
MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - huma  
C:Species: Homo sapiens (man)  
C:Date: 10-Aug-1990 #sequence\_revision 06-Nov-1992 #text\_change 15-Sep-2000  
C:Accession: B35098  
R:Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.  
Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990  
A:Title: A gene pair from the human major histocompatibility complex encodes large pr  
A:Reference number: A35098; MUID:90192810  
A:Accession: B35098  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2142 <BAN>  
A:Cross-references: GB:M33509; NID:q179338; PIDN:AAA35585.1; PID:q179339; GB:M31293  
A:Note: the authors translated the codon AGT for residue 97 as Gly

Query Match 2.8%; Score 353; DB 2; Length 2142;  
Best Local Similarity 20.4%; Pred. No. 3.4e-06;  
Matches 439; Conservative 224; Mismatches 833; Indels 656; Gaps 104;

Qy 503 QOIARPS--QEEKVKEEDKAEKTEKKEEKKDEEKD-----EKEDSKENTKEKDKI 554  
Db 299 EPVGRPSLKEDNL-----KEFDQDQENDDQWAGAHEEVDYTEKLKFSDEE 345  
Qy 555 DCTAETE-----EREQATPRGRKTANSQGRKGRITRS--MTNEAAAAASAAAAATE--- 605  
Db 346 DGRDDEGAEGHRDSQSASGEERPEADGKAGNSPNSPEPTPKTAWAETSRPPEPEGP 405  
Qy 606 -EPPPPLPP-----PPEPSTPEVETSRWTEEMEVAKKGLVEHGRNMAAIK 652  
Db 406 PAPKPPLPGDYDRGGPPCKPPAPEDEDEAWRQRKQSSSEISL--AVERAR----- 456  
Qy 653 MVGTSEACQKFNFYNYKRRHNLNLQHKOKTSRKPREERDVSCQESVASTVSAQDE 712  
Db 457 -----RRREERRRQEEERRAACAKL-----KRLDE 483  
Qy 713 DIEASNEENPEDSEVAVKPSDESPENATSR---GNTEPAVELEPTTETAPSTSPSLAV 769  
Db 484 KEGADPKRLKAEPAAPPA--PSTPAPPVAVPKELPAPPAPPPASAPPTETEPE-EPAQAP 541  
Qy 770 P--STKPAEDESVEQVFN-----DSISAEPTAQMDVDQOEHSAE--EGSVCDPPPT--- 817  
Db 542 PAQSTPTPGVAAAPTSLVSGGSGSTSSSGSFEASVPEQLPSKEGPEPEEVPPTTPPV 601  
Qy 818 -----KASDVDEVRVYPENHASKVGEDNTEKRDLDRASEKVEPRDLDLVAAQIINARPE 872  
Db 602 PKVEPKGIGIGTRQPP-----SQGLGPKYQKSL---PPRFQROOQEQLLKQO---QOHO 651  
Qy 873 PQSDNDSSATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPL--KPNPLDLPLQLO 931  
Db 652 WOHQOQGSAPPTVPVPSPPQVTLGAVPAQAPP---PPPKALYPGALGRPPMPMPMNF 708  
Qy 932 HRAAVIPMVWVSCPCNIPITGTPVSGYALYQRHKAHESALLEEORQ-----RQEQIDLE 986











A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 2.78; Score 344.5; DB 2; Length 3507;  
Best Local Similarity 10.38; Pred. No. 1.4e-05;  
Matches 432; Conservative 296; Mismatches 912; Indels 719; Gaps 93;

```
Qy 515 BEKEDKAEKTEKKEEKKDEEKKEDSKENTKEKDKIDGTAEETEERQATPRGRKT 574
Db 761 EDGEPEATATKPCISADPKTEGSKTESGEEKLIVKED----GKAAOSSGSSATSSGKKS 816
Qy 575 ANSOGRRKGRITRMTNPAASAAAA----- 601
Db 817 BATSGSSSSSAKSGTGPASGSSGASSSSGSGSVSTESGSPGTSSGSGVSEA 876
Qy 602 -----ANTEPPPP-----PL-----PPPEPISTPEVETS--RW 628
Db 877 TGGTGVDSGSGKPKSKSTKEKLPFTKNGEKSPISGSDTTKESSEFTSRKPIEGSDSL 936
Qy 629 TE-----BEMEVAKKGLVEHGRNWAALAKMVGTKSEAOCKNFYENKRRHLDNLQOHK 683
Db 937 TEGSGSGEMFETGSKGHESSKSVTSCKGPTQSGAEGSGSKPKVPGKAPCAPEITDGE 996
Qy 684 QKTSKKPREE-----RDVQCESVAST-----VSAQEDIEDIASNEENPE----- 724
Db 997 ESSSTSTGDKSGGKPADKSDKNKVPKTDGKNPDITTDGEDSTSETSGGEGQPKGSKGQP 1056
Qy 725 --DSEVEAVKPE--DSPENAT--SKNTEPAVELEPT-----TTAPSTSPSLAVP 770
Db 1057 PGDKGSEVKKPTSEVDGPGNLSGTGKNSN--VPLKPTDLPPEGSGILTSSGGKNSTFEH 1114
Qy 771 STK-----PAEDSEVET--QVNDISISA--ETAEQMDVDQOHSAAEGSVCDPPPATKADSV 822
Db 1115 GTKLERLPKPKTESSTPQLGLEISAKKPEPEDGTSKEVGLLEILWESITPGSTILDSD 1174
Qy 823 DVEVRVPENHASK-----VEGNTKERDLDRASEKVEPRDEDLVVAQOINAQRPPEPQ 874
Db 1175 SVGLEISGDLTKATKKPHVEIEGSGTGDEITATTDRVSK-----STKKPRVE 1223
Qy 875 SDNDSSATCSADEVDGEPERQRMFMDKPSLLNPTGSLVLSPLKPNLDLPQLQHRA 934
Db 1224 VDGDNGETSG---VDGKPTTPAPTSSSAES-----STSRIPPTSEAP----- 1265
Qy 935 AVIPPMVSCPCNIPIGRPVSGYALYORHIKAMHESALLEBQRQRQEQIDLECRSSTSPC 994
Db 1266 -----EGSGSEA 1272
Qy 995 GTSKSPNREWEVLOPAPHQLITNLPPEGVRLPTTRTPRPPPLIPSSKTTVASEKPSFTMG 1054
Db 1273 GVPESPDGSGE-----SSTSPADGVSP--PTSSATAPEV-----TTSASSTPDAAVEE 1317
Qy 1055 GSIQGTPTGTYLTHNQASYTQETPKPSVSGISILGLPQOESAKSAT--LPYIKQEE--FS 1111
Db 1318 SGIPSTSKPAELETAPSTEVTPSGSGTEESTLPTPEGSGESTSSSAPTVEPATVLP 1377
Qy 1112 PRSONSOPEGLLVRAQHEGVYRGTAGAIQ--EGSI--TRGPTTSKISVESIPSLRGSIQ 1167
Db 1378 PONRNEKPE---PTKDTFALPTTTTGAQANDSSVENTKCTSSDECGLDACCERRTGVC 1434
Qy 1168 GTPALPQTGIPTEALVKGISRMPIEDSSPEKG---REAAASKGHVIYEGK--SGHILSY 1222
```

```
Db 1435 CEPG-----FEGAPPKKSCVDVDECATGDHNCHE SARCONYVG 1474
Qy 1223 -----DN1KNAREG----- 1231
Db 1475 ACFCPTGFRKADQSCQDIDECTEHNSTCCGANAKVKNKPGTYSCEBNGFLGQYQVCP 1534
Qy 1232 -TRSPRTAHEISLRSYSE-----VEGNIK-----QGSMSRESPPVSALEGLICRALP 1278
Db 1535 TTKPKCDSTQSSKSHCSSESNNMCEVDFVDGSEVECKGCGYKSKGVCEDINECVAERAP 1594
Qy 1279 -----RGS-----PHSDLKERTVLSGSIM-- 1297
Db 1595 CSLNANCVMNMGTFSCSCQKGYRGDFGMCTDINECDERHPCHPHA---ECTNLEGSFKE 1651
Qy 1298 -----QGTPRATT---ESPEDGLKYPKQIKRES--PPIRAFEAGAI----- 1332
Db 1552 CHSGFEGDGIKKCNPLERSCEDVEKFCGRVDHVNSCLSVRIYNGSLSSVCBCEPGRFEK 1711
Qy 1333 -----TKG-----KPYDG-----ITTIKEMGR----- 1349
Db 1712 ESNSCVDIDEEESNRNCDPASAVCVNTEGYSRCECARGYEGEGVCTDIDECDRGMAGC 1771
Qy 1350 -----SHEIPROD--ILTOESRKTPEVQSTRPII 1378
Db 1772 DSMAMCINRMGSCGCKMAGYTGDAATCIKIEEPKSKDTACTDEWSRLCELEKKQCTVD 1831
Qy 1379 EGSISQ-----GTPIKFDNNSGSAIKHNKVSILITPCKSLSRGMPLEIVP----- 1424
Db 1832 EEEVPOGACLPGHHPI---NGTCQSL---OISGLCAQKNDCNKHAECIDILHDSHFCS 1885
Qy 1425 -----ENIK-----VVERGKY-- 1435
Db 1886 PDGFIGDMICDDVDECNNAGMCDDENTKCENTIGSFNCVLEGFKKVKDEKCVVDEKQP 1945
Qy 1436 --EDVKAETVRSRHTS-----VYSGSPVLRSLTHEAPKAQLSPGYIDTTSARRT 1484
Db 1946 NREKIEDEENSSSSNSGQEKPTTKGIVSSTSATSSSETTAEPHVTTTSSISSTTKDMS 2005
Qy 1485 PVSQON--TMSGSPMMNRTSDVTIPPNKSTNHERKSLTTPQRESIPAKSPVPGVDVPS 1543
Db 2006 SKSPENVTMSSEP-----EVTSSSSKSTT--ASETVTSVSSPSSSEAPLTSPPATT 2058
Qy 1544 HSPDPDHRHGTAGEVYWSHLPTQDLPAMPFPHRALDPAAYLFQRQLSPPTPGPYSQYL 1603
Db 2059 EVITESSVKSTTPKESSSEITVKLSKSP-----EVTESSVKSSPSTPTTSQ 2107
Qy 1604 YAMENTQTLINDYITSOQOVNLRPDVARGLSPREQLGLPYPATRGIDILTNWPPITL 1663
Db 2108 SVTSTVPETSKSTVLSSEAPVTSPTSEVHTSS--ETKPSLSASSTTGTNTSTT--PSTSS 2164
Qy 1664 VPHPGGTSTPPMDRITYIPGQITFPPRPYNASMSPCPHPLAAASAAEREREREKE 1723
Db 2165 LASVKTSAP-----EGTSASVAPVKUSSLSPDVSPQSPSTKTFDAT-----E 2205
Qy 1724 RERERIAAASDLYLRPGSEQPGRPGSHG---YVRSPSPSVRTOETMLQORPSVFOQTNG 1780
Db 2206 SSTVQASSTSGTSVKSTSE---PESHVTKLSITSNPSSVPVTPSPKSTPTVPESTEQ 2261
Qy 1781 TSVITPDLPTAQLRIMPLPAGGPSISOGLPASRYNTAADALAAALVDAASAP--OMDVSKT 1839
Db 2262 PTSTPTSGQS---LTPM---NSNSEVLTTSEPHVLSLSSLPDVSPQSSTTNNLSEST 2313
Qy 1840 KESKHEAARLEENLRSSAAVSE-----QQOLEQKTLVEKRSVOCLYTSS 1895
Db 2314 VETPKTSS---EVSUNSEPSTTEAPTILSPDILSTTTNNLSQSSTVSTEDRS--EISENS 2370
Qy 1886 AFPSGKQPHSSVVYSEAGDKGPPPKSRYEEELRKTITTAANFIDVITRIQIASDK 1945
Db 2371 EKPTSAPELVTSSTVHVASSSPDVPTESSEPDULTGSSTENIPEASSKQTISSPTPTDT 2430
Qy 1946 DAREGQSDDSSSLSS-----SHRYETPSDATEVLISPASSAPPQEKLOTYQ---PE 1995
Db 2431 TASEPTKSTMSPDLSLTGNNVLSSESTTTPSSSK---SPVSSSTEGISVVTISTEFSKYPE 2488
```







Qy	1813	RYNTAADALAAVLVDAAA---SAPQMD-----VSVTKSKHEA-----ARLEENLRSSRAVVS	1861
Db	2656	RPEDAGEKEHVIKSAVGHKNPEKLDNCHSVSRVKAQGDLSLEAQLSSLESSRRVHTSTPS	2715
Qy	1862	EQOOLEQKTLLEVK-----RSVQCLYTSASPSPG-----KQPHSSVYVYSEACKDKGP	1909
Db	2716	DKNLLDPTYNAEILKSDSDNNNSDDC---GNILPSDIMDFVLKNTFSMAL-----GE	2764
Qy	1910	PKRSYEELRLTRGKTTITAAANFDVILITROASDKDARERGSSQSDSSSLSSSHRYETP	1969
Db	2765	SPSS--SSELLTIGEGGLDLSNREKDIGLFEVFSQQLPATEPVDSVSSVSSISAEOFLP	2823
Qy	1970	SDA---IEVISPASSPAPQ-----EKLTQYQPEVVKVKAQENDPTRQYEGPL	2014
Db	2824	LELPDLSLVLTTRSPVPSQNSRLAVISDSGEKRVITEKSVAS--SEGDPALLSPG--	2879
Qy	2015	HHYRPOQESPSQQLPPPSQAQEMGQVPRTHRLITLADHI-----CQII-----TDQFA	2064
Db	2880	-----VDPAPRGHMTPDHFIQGHMD-----ADHISPPCGSVQEGCHGNSQDLT	2922
Qy	2065	RNQ-----VSSQTPQO-----PPTSTFQNSPS-----ALVSTPVRTKTSNR-----	2100
Db	2923	RNSGTPLQVPVSPVTPVQNKYVPSSDPSGQSLNSNAAVQTTTPHLKPATEKLIIVNQ	2982
Qy	2101	-----YSPESAQSVHHQ-----RPGSRVSP	2121
Db	2983	NMQPLYVLQTLPNQVYQKIQLTSPVSYSTPSVMTNTSVLPGMSGCLTLTTGLNPSLPPSP	3042
Qy	2122	ENLVDSKRGSRPGKSPERSHVSEYEPIS--PPQVVPVHEKQDSLILLSORGAP----	2175
Db	3043	SLFPPASKGILL--SVPHHQHLHSFPAQAQSSFPN-----ISSPPSGLLIGVQPPDPQLLG	3097
Qy	2176	--AEORDARSPGISYLPSPFTKLENTSPMKVSKQKEIFRKLNSGGGDSMAAAQ----	2230
Db	3098	SEANQRTDLTTVA-----TPSSGLUKKRPISRLTRKNKKLAPSSAPNTAPSDVVSNTFLI	3154
Qy	2231	--PQTEIFNLPAVTTSGVSSSRGHSFADPASNLGLEDIIRKALMGSDFKDKVEDHGVVMSQ	2288
Db	3155	NFTPSQLSNHPSLLDLGSL-----NPSHRTVPNIIRKSKSGIM--YFEQAPLLPPQ	3204
Qy	2289	PMGVVPGT-ANTSVVTSGETRREEDPPSHSGVGCKPKLIISKNSR--KSKSPIDG----	2341
Db	3205	SVGCTAATAAGSSTISQDTSHLTSGPVSALASGSSVLNVVSMQTAAPTSTSVPGHVTL	3264
Qy	2342	--QGYLGTERPSSVSVHSEGDYHRTPGWAMEDPSS--TGSTQFPYNPILTMRLMSTP	2397
Db	3265	ANQRLGTPDIGISHLILKASH--OSLG--IQDQPVALLPSSGMFPQ-----LGTS	3312
Qy	2398	PTPIACAPSAVN 2409	
Db	3313	QTPSAAAWTAAS 3324	
RESULT	39		
S33124			
tpr	protein - human		
N;	Alternate names: kinase-related transforming protein (tpr-net); protein w		
C;	Species: Homo sapiens (man)		
C;	Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-2000		
C;	Accession: S33124; S23740; S00928; G01185		
R;	Mitchell, P.J.; Cooper, C.S.		
Oncogene	7, 2329-2333, 1992		
A;	Title: The human tpr gene encodes a protein of 2094 amino acids that has		
A;	Reference number: S33124; MUID:93064711		
A;	Accession: S33124		
A;	Status: preliminary; nucleic acid sequence not shown; translation not sho		
A;	Molecule type: mRNA		
A;	Residues: 1-2094 <MIT>		
A;	Cross-references: EMBL:X66397; NID:g633225		
A;	Note: the nucleotide sequence was submitted to the EMBL Data Library, Oct		
R;	Mitchell, P.J.; Cooper, C.S.		
Oncogene	7, 383-388, 1992		
A;	Title: Nucleotide sequence analysis of human tpr cDNA clones.		











QY 1479 TSARTPVSYQNTMSRSGPMNRTSDVT---IPPNKSTNHKST-LPTQRESIPAKSP 1534  
DB 1159 SLP-----EGATSPGPRREARPPQVCGWSPPAKSLAPKPTGTLPLPPSKPEKEL- 1210  
QY 1535 VPGVDVPSHSPFDPHHRGSTAGEVYWHLTQLPDPAPEFHRALDPAALYLFQRLSPT 1594  
DB 1211 IFG-----PLSPVARGSGNG---GNGVME-DGERPRRRHGRA-----QQQ----- 1248  
QY 1595 PGYPSQYQYAMENTROTILNDYITTSQOMQVNLPRDVAAGLSRPRQPLGLPYPATRGIID 1654  
DB 1249 -DKPPFRRLKOE-----RENAARG-SEGKPSLTLPASA----- 1280  
QY 1655 LTNMPTTLVPHPGTSTPPMDRITYIGTQITPPRPYNASMSGP-----CHPTH 1705  
DB 1281 -----PG-----PEAL-----TTVTAPAPRAAAKSPDLNSQNSDQANBEW 1318  
QY 1706 LAAASAEREREREKE-----LPAARYNTAADALAAALVDAASAPQMDVSKTKESKHEAA 1847  
DB 1319 ETASESDFTSERRGDKKAPPPVLLTPKAVTPGGGGGAVPGISAMSRGDLUSQRAKDL 1378  
QY 1738 LRPGEQPGRFGSHGYVRSPSPSVRTQETMLQQRPSPVQGTNGTSVITPDPQALRIMP 1797  
DB 1379 KRFSFQQ---RPMERQNRPPGPGGA-----GSSGSS-----SG 1410  
QY 1798 LPAGGSPISQ-----LPASRYNTAADALAAALVDAASAPQMDVSKTKESKHEAA 1847  
DB 1411 GAGGPGGRTGPRGDKRSWSPKNSRPPPE-----RPPGLPLPPPPSSSAVF 1460  
QY 1848 RLEENLRSAVSAVQOOLEKTLVEKRSVQCLYTSSAFPSGKPPQPHSSVVSAGDK 1907  
DB 1461 RLDQVHINPAGI---QQALAO---LSSRQGSVT---APGGHPRHKPGP-----POAPQGP 1507  
QY 1908 GPPPKRYEELRTGKTTITAAANFIDVITRQIASDKADARERGSSQSDSSSLSSHYE 1967  
DB 1508 SPRPTRYEPQVRNSG-----LSSDFHFEPPGVPVRGVG----- 1542  
QY 1968 TPSDAIEVSPASSPAPQOEKLO-TYOPEVVKANOENDPTRQYEGPLHHVRPOQESPP 2026  
DB 1543 TPRDSAGV-----SPFPKRERRPPRKPPELL---QEESLPPPHSSGFL---GSKPEGPGP 1591  
QY 2027 QOOLPPSSQAEQMGQVPRTHRLTLADHICQIITQDFARN---QVSSOTPOQPPSTFQ--- 2082  
DB 1592 QAE-----SDTQTEALTPHWNRLHATSKYRPSMSMEPMEPLSPFEDV 1638  
QY 2083 -----NSPSALVSTPVRTKTSNRYSPESQAQSVHH---ORPGSRVSPENLV 2125  
DB 1639 AGTEMSQSDSGDLSDGQVSSGP-----CSQSSPDGGLKGAEGPPRPGGS-SPLNAV 1693  
QY 2126 DKRSRPGKSPERSHVSSEPEYIPSPQVPVVEKODSLLLLSQR-----GAEPABQR 2179  
DB 1694 -----PCEGPP-----GSEP-PRRPPAP-HDGDRLKELPREQPLPPPIGTERSQRT 1738  
QY 2180 NDARSPGISVLPSTFFTKLENTSPWK-----SKQEIFRKLNSGGGSDMAA-----AQP 2231  
DB 1739 DRATEPGPIR-PS-----HRGPPVQFGTSDKQSDSLVVGVDLKAELKELTASVTEAIP 1791  
QY 2232 GTEIFN-LPAVTTSGVSSR-----GHSFADPASNLGLIEDIIRKALMGSDFKDVEDHGVVM 2286  
DB 1792 VSRDWELLPSAAAEQSKNLDGSHCVPEPSSS---GORLYPEVFGS-----AGPSS 1842  
QY 2287 SOPMGVPGTANTSVTSGETRRREGDPSPH 2317  
DB 1843 SQ-----ISGGSHGLSITSQWRLRPGTSLH 1869

## RESULT 41

S13507

microtubule-associated protein MAP2 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999

C:Accession: S13507

R:Marchal, D.; Delapierre, D.; Dresse, A.

Arch. Int. Physiol. Biochim. 96, 231-236, 1988

A:Title: Cloning and partial sequencing of a new rat brain specific cDNA.  
A:Reference number: S13507; MUID:89334524  
A:Accession: S13507  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-1825 <MAR>  
A:Cross-references: EMBL:X54100; NID:g56624; PIDN:CAA38034.1; PID:g56625  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1990  
C:Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology  
C:Keywords: microtubule binding; tandem repeat  
F:1667-1697/Domain: MAP2/tau repeat homology <MT1>  
F:1698-1728/Domain: MAP2/tau repeat homology <MT2>  
F:1729-1760/Domain: MAP2/tau repeat homology <MT3>

Query Match 2.7%; Score 338; DB 2; Length 1825;

Best Local Similarity 20.1%; Pred. No. 1.1e-05;

Matches 359; Conservative 232; Mismatches 587; Indels 606; Gaps 94;

QY 1008 QPAPHQITNLPEGVRLPTTRTPRPPLIPSSKTTTVAASEKPSFIMGSGISQGTPTGLT 1067  
DB 119 QPAALPLAAE-ETVNL-PSPPSPASEQTAALAEAS----- 153  
QY 1068 SHNQASYTOETPKPSVGSISLGLPRQESAKSATLPYIKQE-EFSPRSONSOPEGLLVRA 1126  
DB 154 -----KMEFPQOKLPSPFABPLDKEETEFRKMQSKPGED-----F 188  
QY 1127 QHEGVV--RGTAGATQ-----EGSITRGTPTSKISVESIPSLRGSITQCTPALPQTG 1176  
DB 189 EHAALVPODTSKTPQDKKPOOMEKSPASFPATQGTNLEDIK-QITEPSITVPSIG 247  
QY 1177 IPTALV---KGSIRMPIDESSPEKREEAASAGHVI-----YEKG----- 1215  
DB 248 LSAEPLAPKQDKQDFIEMPVESKKDENGAAPISGPLTPMREKDVLEDIPWEGKQFDS 307  
QY 1216 -----SGHI-LSYDNINKNAR--EGTRS-----PRTAHEISLKRYSYSE 1251  
DB 308 PMPSPFHSGSETPLDPTVKDERVTEGSPFAPVFFQSDKMSLQDTSGSATSKESKDBE 367  
QY 1252 GNIKQGMRESVPSAPLEGILICALPRGSHSLDKERTVLSGIMOGTTPRATTFESFEG 1311  
DB 368 POKKADKADVDPVSEATVTL-----GDVHSPAVEGFV--GENISGEKGTID-- 413  
QY 1312 LKYPQIKRESPPIRAFEGAITKGPYDITIKEMGR-SIHEIPRODILITQESKRTPE- 1369  
DB 414 -----QEKETSTPSVQPTLTETEPO---TKLETSKVSIETVAKE---EESLKLDD 462  
QY 1370 ---VVQSTRPIEGSISQGTPIKFDNNSGQ-----SAIKH-----NVKSLITGPSKLSRM 1417  
DB 463 KAGVIQTS---TEHSFSK-----EDQKEQETIEALKQDSFPISLEQAVTDAAMATKL 513  
QY 1418 PPLEIVPENI--KVVERGYEDVKAGETVRSRHTSVVSGSPVSLRSTLHEAPKAQLSPGI 1475  
DB 514 EKVTSPEAVSEKREIOGLFEEDIADK-----SKLEGAGSATVAEV--EMPYEDKSGM 565  
QY 1476 --YDTSARRTPVS-----YQNTMSRSGPMNRTSDVTIPPNKSTNHKSKTSLPT 1524  
DB 566 SKYETSALKEDVTRSTGLGSDYELSDSRGN--AQESLDTVSPKNQ---QDEKELLAKA 620  
QY 1525 QRESIPAKSPVGVDPVSVHSPFDPHHRG-STAGEVYVSHLPTQL--DPAMPFHR--ALD 1579  
DB 621 SQSPSPA-----HEAGYSTLAQSYTSDHPSELPEEPSPSSPOERMFTID 662  
QY 1580 PAAAYLFQROLSPTPGYPYQYQLYAMENTROTILNDY-----ITSQQQVNL----R 1628  
DB 663 PKV-----YGERKDLHSHKKNKDDTLRSRLGLGGRSAIEGRSMINPLMSCL 708  
QY 1629 PDVARG-----LSPREQPLGLPYPATRGIIDLTN--MPETILVPHPGTSTPMDRI 1678  
DB 709 DSIALGFNFGRHDLSPLASDI---LTNTSGTMDGDDYLLPT-----TPAVEKI 755  
QY 1679 TYIP-----GTQIT-----FPPRPY--NSASMSGPHPTHLAAA-- 1709







[illegible]

RESULT 43

RESOL  
T45025

14023 mucin MUC5B, tracheobronchial [imported] - human (fragment)

C:Species: Homo sapiens (man)

```
C:\pcceres> nrmc buprens (man)
C:\pcceres> 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
```

C;Accession: T45025

R;Desseyn, J.L.; Guvonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

J. Biol. Chem. 272, 3168-3178, 1997

**A:Title:** Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter

A:Reference number: Z22899; MUID:97166151

A: Accession: T45025

A: Status: preliminary: translated from GB/EMBL/DDBT  
A: Accession: 145025

A-Molecule type: DNA

A; MOLECULE TYPE: DNA  
A: RESIDUES: 1-3570 <DEF>

A;Residues: I-35/0 <DES>

A; Cross-references: EMBL: Z/2496

A; Experiment

C;Genetics:

Query Match 2.78: score 336: DB 2: Length 3570:

Best Local Similarity 18.9%: pred. No. 3e-05

Best Local Similarity 10.9%, F1ed: NO: 38-05,  
Matches 410: Conservative 228: Mismatches 830: Indels 696: Gaps 91:

Qv 567 ATPRGRK TANSOGRRKGRITRSM TNEAAAA SAAAAAATFEPPPPPI.PPPPEPTSTEPVETS 626

QY J07 AIFRGNRIANSQQARRKROKTIASMINETATRASSTATALEEFFEFISIEFVEIS QZ

017 627 B-----WTF-----FEMV-----AKKCY VEHCPNWAATL-KM 653

QY 62/ R-----WTE-----EEMEV-----AKKGLVEHGRNWAATA-KM 653

Case	Year	Country	Population	Age	Gender	Occupation	Education	Income	Health	Family	Social	Environment	Policy	Outcome
1	2010	USA	300M	65	M	Retired	High School	\$25K	Good	2	Low	Urban	Active	Survived
2	2011	USA	300M	65	F	Homemaker	High School	\$15K	Good	3	Low	Urban	Active	Survived
3	2012	USA	300M	65	M	Teacher	College	\$35K	Good	2	Low	Urban	Active	Survived
4	2013	USA	300M	65	F	Nurse	College	\$45K	Good	2	Low	Urban	Active	Survived
5	2014	USA	300M	65	M	Engineer	College	\$65K	Good	2	Low	Urban	Active	Survived
6	2015	USA	300M	65	F	Manager	College	\$75K	Good	2	Low	Urban	Active	Survived
7	2016	USA	300M	65	M	Doctor	College	\$120K	Good	2	Low	Urban	Active	Survived
8	2017	USA	300M	65	F	Lawyer	College	\$110K	Good	2	Low	Urban	Active	Survived
9	2018	USA	300M	65	M	Executive	College	\$130K	Good	2	Low	Urban	Active	Survived
10	2019	USA	300M	65	F	Executive	College	\$140K	Good	2	Low	Urban	Active	Survived
11	2020	USA	300M	65	M	Executive	College	\$150K	Good	2	Low	Urban	Active	Survived
12	2021	USA	300M	65	F	Executive	College	\$160K	Good	2	Low	Urban	Active	Survived
13	2022	USA	300M	65	M	Executive	College	\$170K	Good	2	Low	Urban	Active	Survived
14	2023	USA	300M	65	F	Executive	College	\$180K	Good	2	Low	Urban	Active	Survived
15	2024	USA	300M	65	M	Executive	College	\$190K	Good	2	Low	Urban	Active	Survived
16	2025	USA	300M	65	F	Executive	College	\$200K	Good	2	Low	Urban	Active	Survived
17	2026	USA	300M	65	M	Executive	College	\$210K	Good	2	Low	Urban	Active	Survived
18	2027	USA	300M	65	F	Executive	College	\$220K	Good	2	Low	Urban	Active	Survived
19	2028	USA	300M	65	M	Executive	College	\$230K	Good	2	Low	Urban	Active	Survived
20	2029	USA	300M	65	F	Executive	College	\$240K	Good	2	Low	Urban	Active	Survived
21	2030	USA	300M	65	M	Executive	College	\$250K	Good	2	Low	Urban	Active	Survived
22	2031	USA	300M	65	F	Executive	College	\$260K	Good	2	Low	Urban	Active	Survived
23	2032	USA	300M	65	M	Executive	College	\$270K	Good	2	Low	Urban	Active	Survived
24	2033	USA	300M	65	F	Executive	College	\$280K	Good	2	Low	Urban	Active	Survived
25	2034	USA	300M	65	M	Executive	College	\$290K	Good	2	Low	Urban	Active	Survived
26	2035	USA	300M	65	F	Executive	College	\$300K	Good	2	Low	Urban	Active	Survived
27	2036	USA	300M	65	M	Executive	College	\$310K	Good	2	Low	Urban	Active	Survived
28	2037	USA	300M	65	F	Executive	College	\$320K	Good	2	Low	Urban	Active	Survived
29	2038	USA	300M	65	M	Executive	College	\$330K	Good	2	Low	Urban	Active	Survived
30	2039	USA	300M	65	F	Executive	College	\$340K	Good	2	Low	Urban	Active	Survived
31	2040	USA	300M	65	M	Executive	College	\$350K	Good	2	Low	Urban	Active	Survived
32	2041	USA	300M	65	F	Executive	College	\$360K	Good	2	Low	Urban	Active	Survived
33	2042	USA	300M											

QY 654 VG-----TKSEAQCK-----NFYFNYKRR-----HNLDNLLQHQKQTSRKPR 691

Qy 692 EERDVQCESVASTVSAQEDEDIASNEENPEDSEVE-----AVKPSEDSPENATSRG 745



Qy 746 NTEPAVELEPTETAPS-----TSP-----SLAVSTKPAEDESVEY 782  
Db 638 TALPALRSTATPTATSTFAIPSSSLGTTWRLSQTTPMATMSTATSSP---ETVHT 694  
Qy 783 QVNDISIAETAEQMDVDQOEHSABEGSCVDP--PPATKADSVDEVVRPNHASKVEGDN 840  
Db 695 STVLTTTATT-----TGATGVSATPSSTPGTA-----HTTKVLTTT 730  
Qy 841 TKERDLDRASEKVEPRDEDLVVAQINAEQPEQSDNDSSATCSADEVDGCEPERQMF 900  
Db 731 T-----TGFTATPSS-----SPGRARTLP 749  
Qy 901 MDSKPSLLNPTGSLVSSLPKP-----NPLDLPOLAHRAAVIPPMVSCTPCNIPIG-- 951  
Db 750 -----VWISHTTPTTRGSRVTPSSIPGTH-----IPTVLTTTITIVATGSM 792  
Qy 952 -TPVSGYALYORHIAKHESALLEEQRORQOIDLCEKSSPSGCTSKSPNREVEVLOPA 1010  
Db 793 ATP-----SSSTQTSCT----- 804  
Qy 1011 PHOLITNLPE-----GVRLPTRP--TRPPPLIPSSKTTVASEK-----PSFIMGSSISQ 1060  
Db 805 PPSLTATTATGATGTTNPSSTPGTTPPVLTTTATPAATSTSTVTPSSALGTTHTPP 864  
Qy 1061 TPGVYLTSHNOA-----SYTOET-----PKPSVGSISLGLPROQESAKSA 1100  
Db 865 VPNTTATTHGRSLSPSSPHTVCTAWTSATSGILGTHITEPSTGUSH--TPAATGTGTQH 922  
Qy 1101 TLPYIKQEFSPRSONSOPEGLLVRAQHEGVVRGAGAIQSGSTRG--TPTSKISVESI 1158  
Db 923 STPALSSPHSPSSRTESP-----SPGTTTPGHTTATSTRTATAT 962  
Qy 1159 PSLRGSITQGTALPQGTGIPTEALVKGSIRMPIEDSSPEKREAAKGVHVEGRSGH 1218  
Db 963 PSK-----FRTSLTPSQ--PISA-----PITTVTMGCEPQCAWSEWLDYSYMPGPGSGD 1012  
Qy 1219 ILSYDNINKAREGRSPRTAHEISLKRYSVEGNIKOGMSMRE--SPVSAPLE-GLICR 1275  
Db 1013 FDTYSNIRAAGGAVCEQPLGLE-----CRAQAQGVPLRELQGVVECSLDFGLVCR 1063  
Qy 1276 -----ALPRGSPHDLKERTVLGSSIMOGT 1300  
Db 1064 NREQVGKFKMCFNVEIRVFCNNGHCPSTPATSTATSSSTPGT-----TWILTTLT 1115  
Qy 1301 PRATTESPEDGLKVPKQIKRESPPTR-----AFEGAITKGK-----PYDGI 1342  
Db 1116 TTATTTSTGTAFTSLTRAPPKPVLTATPTVTSKATPSSSPGCTATPALRST 1175  
Qy 1343 TIKEMGRSIEIPRODILTQESR-----KTPVEVQSTRPIIEGSSISQ 1385  
Db 1176 ATPTATSVTPIPSSSLGTTWRLSQTTPATMTATPSSSTPETAHTSTVLTAATATTG 1235  
Qy 1386 TPIKFDNNNGOSATKHNVK-----SLITGSKLSRGMPLLEIVPENIKVVERKYEVDVAGE 1442  
Db 1236 ATGVSATPSSTPGTAHTTKVPTTTTGTATPSSSPGALTTP-----VWISITT 1285  
Qy 1443 TVRSRHTSVWSSGSPVLKRLSTLHEAPKAQLSPGIYDDTSARTPVSYONTMSRSGSPMMNRT 1502  
Db 1286 TPTTRGSRVTV--PSSIPGTHTATVLTITTTTV-ATGSMATPSS--STQTSGLTPPSLT 1339  
Qy 1503 SDVTIPPKNKSNHKKSTLTPTQRESIPAKSPVPGVDVPSVSHSPDPHHRGSTAGEVYWS 1562  
Db 1340 TATTITATGST-----TNPSSTPGTRIP--PVLTTTATTPAATSSSTVTPS--S 1384  
Qy 1563 HLPQLDPAHFHRAIDPAAAYLFQRLSPTPGYPSQYQIYAMENRQTILNDYITISQ 1622  
Db 1385 ALGTHTPPV-----PNTTATTHGRSLSPSS--PHTVRTAWTSATSGTLGTHITEPS 1435  
Qy 1623 MQVNLRPDVARGLSPREP-LGLPYPATRGIIDLNTNPPITLVPHPGGSPPMDRITYI 1681  
Db 1436 TQTSHTPAATGTQHTSPALSSPHSPSR-----TTESPPS-----PGTTPP--CHTAT 1483  
Qy 1682 PGQTITPPRPYNSASMSGHPHTHLAAASAEERERERERERERERERERERERERERER 1741

Db 1484 SRTTATATPSKTRTSTLLPSSPTSAPIITVVVTMGCEPQ-----CAMSE-WLDYS 1531  
Qy 1742 SEQGRPG-----SHGYVRSPSPSVRTQ-----ETMLQORPSPVFGTNGTSVITPLD----- 1788  
Db 1532 YMPGPGSGGDFDYSINRAAGGAVCEQPLGLECRAQAQGVPLRELQGVVECSLDFGLVC 1591  
Qy 1789 -----PTAOLRIM-----PLPAGGPSISQGLPAS-----RYNTAADAALA 1822  
Db 1592 RNRQVGKFKMCFNVEIRVFCNNGHCPSTPATSTATPSSTPGTTWLTTEQTTAATTTA 1651  
Qy 1823 ALVDAA-----ASAPQMDVSKTKESKHEAARLEENLRSAAYSEQQLEBKQKLEVEKR 1876  
Db 1652 TTGSTAIPSSPTPGTAPPKVLTSQATPTA-----TSSKATSSSPSRATTLPLV-- 1700  
Qy 1877 SVQCLYTSSAPPSKPKQPHSHSVVYSEAGKDKGPPPKSRYEELRTRGKTTTAAFNFDVI 1936  
Db 1701 -LTSTATKSTATSFPTIPSSTLGTTGTGSRPPHPMAT-----MSTLHPSSTPETT 1750  
Qy 1937 ITROIADKDAEREGSOSDSSSSLSLHRY-----ETPSDAIEVISPASSAPP-- 1985  
Db 1751 HTSVLTTKATTTATSSMSTPSTPGTTWLTTELTTAATTTAALPHGTPSPSTGTTWIL 1810  
Qy 1986 QEKLTQTYQEVYKANAQAEENDPTROYEGLHHYRPOQESPSP--OOQLPSSO----- 2035  
Db 1811 TEPSTTATVTPTGSTATASSTRATAGTLKVLSTATTPTVISSKATPSSSPGATALPA 1870  
Qy 2036 -----AEGMCQVPRTHRLIHLADHICQIIHQDFARNQVS--SOTPOQPTSTFQNSP 2085  
Db 1871 LRSTATPTATSVTAIP--SSSLGTAWRLSQTTPPTATMTATPSTSTPETVHTST-- 1924  
Qy 2086 SALYSTPVRTKTSNRYSPESQAQSVHHOR-----PGSRVSP-----ENL 2124  
Db 1925 -VLTTTATTTTIGSVATPSSPTPGTAHTTKVPTTTTGTATPSSSPGALTTPPVWISTTT 1983  
Qy 2125 VDKSRGS--RPGKSPERSHVSSEPEYEPISPPQVPVHVEKQDLSLLLSQSGAEPAEQRNDA 2182  
Db 1984 TPTTRGSRVTPSSIPGTHAT-----VLTTTITTTVA----- 2015  
Qy 2183 RSPGISYLPSPFFKLENTSPWVSKKQOEIPRKLNSGGGSDMAAQAOPGTEINLPAVT 2242  
Db 2016 --TGSMATGSSSTQTSQTPSLTTTATTI-----TATGTTNPSTPPTPIP-PVLT 2065  
Qy 2243 TSGSVSRGHSFADPASNGLIEDIIRKALMGSDFKVEDHGVMSQPMGVVPGTANTSVV 2302  
Db 2066 TTATTPATSTSTVTPSSALGTH-----TPPVNTTAT 2098  
Qy 2303 TSGTRREEDGPPSPHSGVCKPKLISKNSRKSPIPGQCYLGTERRPSSSVSHSEGDY 2362  
Db 2099 THG---RSLPPSPHT---VPTAWTSATS-----GILGTHITEPST---GTS 2137  
Qy 2363 HROTTPGHAWEDRPSTGSTQTPPYNPLTMRMLSSPTPIACAPSAVNAQAAPHOONRIWER 2422  
Db 2138 H--TPA-----ATTGTTQ-----PSTPAL-----SSPHSPSRSTES 2166  
Qy 2423 EPAP 2426  
Db 2167 PPSP 2170

## RESULT 44

T03454

ALR protein - human

C;Species: Homo sapiens (man)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

C;Accession: T03454

R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya  
Oncogene 15, 549-560, 1997A;Title: Structure and expression pattern of human ALR, a novel gene with strong homo  
A;Reference number: Z14954; MUID:97388474

A;Accession: T03454

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA



A:Residues: 1-5262 <PRA>  
A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285  
C:Genetics:  
A:Gene: ALR  
A:Map position: 12  
C:Superfamily: human ALR protein  
C:Keywords: alternative splicing

Query Match 2.64; Score 335; DB 2; Length 5262;  
Best Local Similarity 19.48; Pred. No. 5.3e-05;  
Matches 466; Conservative 267; Mismatches 882; Indels 786; Gaps 113;  
QY 566 QATPRGRKANSQGRKGRITRSMTEAAAAASAAA-----AAATEE-----ppp---p 610  
DB 393 QGQPKGGHVTSMQKEPG-----PQCEAKPLGKAGVQLEQLAEPLNEEMPLPPPEESP 448  
QY 611 LPPPEPEISTE--PVTSRWTEEMEVAKGLVEHGRNNAIAKMWGTTKSEAQCCKNFYNY 669  
DB 449 LSPPEESPTSPPEASRLSPPEELPASLPE-----AL----- 483  
QY 670 KRRHNLNLLQHKQKTSRKPRERDVQCSVASVSAQEDIEDIASENEENPEDSEVE 729  
DB 484 -----HLSPLEESPLSPPEESPLS-----PPPESSPFPLEESP----- 519  
QY 730 AVKPSSEDPENATSRGNTPEAVE---LEPTTETAPTSPSLAVPSTKPAEDESVEQTQND 786  
DB 520 -LSPEESPPSPALETPLSPPEASPLSPPEESPLSPPEELPSPPEASRLSPPEE 578  
QY 787 SISATASOMDVQOEHSAE-----EGVCDPPPAKADSDVDVVRVFNHASKY----- 836  
DB 579 SPMSPPEESPMSPPEASRLSPPEESPLSPPEESPLS-----PPPEASRLSPPEE 631  
QY 837 -----EGDNTERDLDRAS-----EKVEPRDELVAQINQAQPE-----PQSD 876  
DB 632 DSPMSPPEESPMSPPEESRLSPVVSRLSPPEESPLSPPEESPLSPPEELPPEFGAKGD 691  
QY 877 NDSSATCSA---DEDVDGEPRQMFMDSKPSLNLPT--GS--ILVSSPLKPNPLD---LP 928  
DB 692 SDPESPLAAPILETIPSPPEANCTDPEVPPMILPPSPGSPVGPASPLMEPLPPQCS 751  
QY 929 QLOHRAVIP---PWSCTPCNIPIGTVPVGYALYQRIHKAHESALLEQRQOEIDL 985  
DB 752 LLOH--SLVPQNSPQSPALPLSV--SPLSPGKGVGVSDAEALHEMETEYVS---EP 806  
QY 986 ECRSSTPCGTSKSPNREWEVLQAPHQL-----ITNLPE-----GYRLPTRTTRPP 1034  
DB 807 EC--PALEFSATSPSPMGDLSCFAPSAPALDDFSGLGEDTAPLDGIDAP---GSQPEP 862  
QY 1035 PLIPSSKTTVASEKPSFTMGGSISQGTGTLYTSHNQAS-----YTQETPKPSV----- 1083  
DB 863 GQTPGSLASELKGSPVLLDPEELAPVTMEVYPECKOTAGRGSPCEOEPRAPVATPP 922  
QY 1084 ----GSISGLPQOESAKSATLYIKOEPRSPRQNSQPEGLLVRAHQEGVVRGTAGAI 1139  
DB 923 TLIKSDIVNEISNLQGDASAFP-----GSEPLLGSPDEGGGSLSMELGVSTDVSPAR 977  
QY 1140 QEGST-----TRCTPT--SKIS-----VESIPSLR 1162  
DB 978 DEGSURLCTDSLPTDSDLLCDAGTAISGGKAEGEKGRRSSPARSRKQGRSSSFPGR 1037  
QY 1163 -----GSITQGTPALPQGTIPTEALVKGISRMPIDSDSPKEGREB-----AASKGHVI 1211  
DB 1038 RPRGAHGRGRARLAKSTASSIETLVADI-----DSSPSKEEEDDDTQNTQVVL 1091  
QY 1212 Y-----EGSGHILSYDNIKNAREGTRSPRTAHEISLKRYSVES 1251  
DB 1092 FSNIDKFLVMQDMCVGSGFRGAEGHLLACSQCQCVHPYCVNSKITKVMLLKGRVCE 1151  
QY 1252 GNIKQGHMSRSP-----VSAPLEG-----LICRALPRGSP-- 1282  
DB 1152 CIVCEVGQASDPRLLLCDDCDISYHTYCLIDPPLLATVPKGGWCKWCVCSCMQCGAASPG 1211

QY 1283 -HSD-----LKERTVLGSGSIMQGTTPRATTESF--EDGLKY- 1314  
DB 1212 PHCEWQNSYTHCGPCASLVTCPICHAPVEEDLLILOCHCEKRWMAHGESLETDVDVHA 1271  
QY 1315 -----PKQIKRESPPIRAFEGA-ITRGKPYDGTITTE 1346  
DB 1272 PDEGDCVSCPYVVKVPVAPVAPPELVPMKVKEPEQYFRFEGVWLTE---TGMALLRN 1327  
QY 1347 MGRS-IHE-----IPRQ-----DILTOESRK-----TPFVVQSTRPI--TEG 1380  
DB 1328 LTMSPHLKRRORRGLGUPGEAGLEGSEPSDALGDDDKDGLDTELLKGGEGVEHREC 1387  
QY 1381 SISQCTPLIKFNNNSQSAIKHNK-----SLITGPKS-----LS 1414  
DB 1388 EIKLEGPVSPDVEPKKEETESKKRKRKPYRPGICGFVWRKRSHTRTKGPAAQAEVLS 1447  
QY 1415 RGMPELEIVPENIK---VVERGKYE-DVKAGETVRSRHTSVSS--GPSVL-----RSTL 1463  
DB 1448 GDGQDEVIPADLPAGAVEQSLAEGDEKKQORRGRKRKLEGMPAYLOEAPFGKELL 1507  
QY 1464 HEAPKAQLSPGIYDDTSARRTPVSYQNTMS--RGSMMNRTSD-VTIIPNKSTNHERKST 1520  
DB 1508 DLSRKALFVGVGRPSFGLGTPKAKGSGSERKELPTSQKGGDDGPDIADEESRLEGKAD 1567  
QY 1521 LTPQRESIPAKSPVPGVDVVSHPFHHRGSTAGVYWSHLPTQDLPAMPFHRALDP 1580  
DB 1568 -TPGEDGVKASPVPS-----DPEKPG-TPGE---GMLSSDLDR----- 1603  
QY 1581 AAAAYLFORQLSPTPGYPYOQLYAMEN--TRQTLINDYITSOQMOVNLRPDVGRLSPRE 1639  
DB 1604 -----STEELPKMESKDLQQLFKDVLGSE-----RE 1629  
QY 1640 QPLG-----LPYPATRIIIDLTMPPYI-LVPHPGCTSTPPMDR-----ITY 1680  
DB 1630 QHLGCGTGLGSRTPQLQRFLOGLPLGNLFSSPMDSYFCLQCSPLDFRERGGFSP 1689  
QY 1681 IPGQITPPRPY--NSASMSPGHPTLAAAAAE-----RERERER-EKERERERIAAASS 1734  
DB 1690 EPGE---PDSPTGSGGTTPTTPTTEGEGDGLSYNQSLQKWEKDEELGQLSTISP 1745  
QY 1735 DLYLR---PGSEQGRPGSHGYVRSPSPSVRTQETM----- 1767  
DB 1746 VLYANINPNLQD-----YPMWSSRCQIMKLWRKVPAAADKAPYLQKADNRAA 1795  
QY 1768 -----LQORPSVFOGTNGTSTVITPLD-PTAQLRI-----MPLPAGSPISOGILPA 1811  
DB 1796 HRINKVOKAESQINKQTKVGDIAKTRDPAHLRIKPPQPGALGSPPPAAAAPTIFIGSPT 1855  
QY 1812 SRYNTAADALAAALVDAASAAPQMDVSKTESKHEA-ARLEENLRSRAAVSEQQOLEQKT 1870  
DB 1856 TPAGLSTADGFLAPPAGSVFPGD-----SPGELFLKLPPQVPAQAPSQDPFLGAPAYP 1909  
QY 1871 LEVERKSVOCLYTSAPPSGPK-QPHSSVYVSEAGKQK-----GPPPKSYEEB----- 1918  
DB 1910 LEPRFTAPPTYPYPSPTGAPAPMLGASSRPGAGOPGEFHTTPCPTPRHQPSPTDPF 1969  
QY 1919 LRTGKTIITANAEDVIITQIASDKDARERGSSSSSSLSHRYETFSDALEVISP 1978  
DB 1970 LKPCPS-----LD-----NLAVSPSGVGGKASE--PLLSPPPGESRKALEVKE 2015  
QY 1979 ---ASSPA--PPQ-----EKLQTYQEVVKANQANDPTROYEGPLHHYRPQ 2020  
DB 2016 ELGASSPSYGPNNLGFVDSPPSGHLGLELKTDPDVFKAPLTPRASQVPSQGLGLRPQ 2075  
QY 2021 QESPSQOQLPPS-----SQAEQMGQVPRTHRLITLADHICQIITQDFARNQVS 2069  
DB 2076 E--PPAQAALAPSPPSHPDIFRPGSYTDYPAQPLTPR----- 2111  
QY 2070 SQTQQPPTSTFQNSPALYSTVPTRTKTSNRY-----SPESQ---AQSVHHORPGSRVS-- 2120  
DB 2112 ---PQPPP-----PESCCALPPRLSPDPFSRVPVSPQSSSSQSPPLTPRLPSAEAF 2161  
QY 2121 PENLVDRSGSRPGKSPERSHVSSEPEPISP---POVPVVEKQDSILLILLS-QRGAEP 2176



```
Db 2162 PSPVTPREFQSDPYSRPPSRQSRDPFAPLHKPPRPQPPEVAFKAGSLAHTSLGAGGPPA 2221
QY 2177 -----BQRDARSQGISYILPLSPFFTKLENTSPMVSKKQEIFRKLNS 2219
Db 2222 ALPAGPAGELHAKVPSGQPPNFVRSPGGAFVGT-----PSMRFTFPQ----- 2265
QY 2220 GGGSDMAAAGPGEIFNLPAVTTSGSVSSRG-HSFADPASNLGLIEDIRKAL-MGSP-- 2275
Db 2266 -----AVGPELSLK-----PPVQGLPPHGINSHFGPGTILGKQPOSTNVTATGNFHP 2314
QY 2276 -----DKVEDHGVVMQPGVVP-----GTANTSVVTSGETRREE-- 2311
Db 2315 SGSPILGPGSGTGESYGLSPLRPSPVLPAPPADGSLPYLSHGASQORSIGTSPVEKREDPG 2374
QY 2312 -----GDPSPHSGGVCKPKL-----ISKNSRKS----- 2336
Db 2375 TGMGSSSLAELPQTQDPMGSLGSLQTELEKQORQRRLIRIQIQRNLTURQEKETAAA 2434
QY 2337 --SPIPGGYLGTERRPSSVSVHSEGDYHROTPCWAWEDRPSSTGSTOFPYNPLTMRMLS 2394
Db 2435 AAGAVGPGSGWGAEPSSPAFQLSRG-----QTPAGTODKSLVG---LPPSKLSGPILG 2487
QY 2395 -----STPPTPIACAPSAVN-----QAAPH-----QONRIWEREP 2424
Db 2488 PGSPFDDRLSRPPPPAPPSSMDVNSRLVGGSAFYORAPYPGSLPLOOQOOLWQOQ 2547
QY 2425 A 2425
Db 2548 A 2548
```

## RESULT 45

JC5630

TCOF1 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 14-Oct-1997 #sequence\_revision 07-Nov-1997 #text\_change 05-Nov-1999

C:Accession: JC5630

R:Paznekas, W.A.; Zhang, N.; Gridley, T.; Jabs, E.W.

Biochem. Biophys. Res. Commun. 238, 1-6, 1997

A:Title: Mouse TCOF1 is expressed widely, has motifs conserved in nucleolar phosphoprote

A:Reference number: JC5630; MUID:97445113

A:Accession: JC5630

A&gt;Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1320 &lt;PAZ&gt;

A:Cross-references: DDBJ:AF001794; NID:g2109458; PIDN:AAB71347.1; PID:g2109459

C:Comment: This protein is a nucleolar phosphoprotein with 82 potential phosphorylation

```
Query Match          2.6%  Score 334.5;  DB 2;  Length 1320;
Best Local Similarity 19.9%  Pred. No. 9.7e-06;
Matches 319;  Conservative 197;  Mismatches 610;  Indels 481;  Gaps 70;
```

QY 503 QQIARPSQEEKVEEKEDKAEKTEKKEEKKDEKEDEKEDSKENTKEKIDGTABETE 562

Db 54 QQTSELGQKQAEDETLQAKSRVSDPVSSSSSDQKEE-----EAAT 98

QY 563 EREQATPR-----GR-KTANSQGRKGRITRSMTEAAAAAASAAAAA 603

Db 99 ERKATPRPTPVNSATAALPSKVEKCKTKTAN-----KTVNSVSHPGSGKTVHLLS 151

QY 604 TEEPPPLPPPEPTSTPEVTSRWTEEME-----VAKGLVHGHNMAAI----- 650

Db 152 GKSP-----KKAEPPLANTVLAEE--TEEGNAQALGPTAKSGTSAQSGSSSEDSISS 205

QY 651 ----AKMVGTKSEACKNFYNYK-----RRHNDLNLQHQKTKSRKPREEDVQ 698

Db 206 DETDVEVKSAPAKAQAASAPAKDPAPTAPGPTKLGTV-----APTAPKAPAAAAA 260

QY 699 CESVASTVSAQDEIDIASNEENPEDEVEAVKPSDENATSRGNTEPAVELEPTE 758

Db 1126 SSSPA-PTQILPNSITQRLLEQ-----AWPLSEAQVQA 1157

```
Db 261 AAAVAAAAAAAE--SESEEDSDSEDEAPAGLPQ---VKASGKG---PHVRADSVSA 312
QY 759 TAPS-TSPSLAVP--STKPADES--VETQVNDISIAETAQMDVDQOEHSAEESGVCDDPPP 815
Db 313 KGISCKPILATPGTKTQPAATQAKAERPKDS---EFSSEDDSDSE----- 355
QY 816 ATKADSDVDEVVRPENHAS---KVEGDNTKERD-LDRASEKVEPRDEDLVVAQIQAOR 870
Db 356 ---DEMPVTVNTQARTSGKSPRARGTSAPAKESSQKGAFAVTPGKARPVAAQ---AGK 408
QY 871 PEPOSDNDSDATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNLDLPQL 930
Db 409 PEAKSSEES-----ESDSGE-----TPAAATLTTPSAKVKPLG--- 441
QY 931 OHRAAVTPMVWCTP-----CNIPICTPVSGYALYORHIKAMHESALLBEORQREQID 984
Db 442 --KSSQVRPVSTVTPGSGGKANLPCPGKVGSAALRVQMVK-----KEDVSESSAE 491
QY 985 LECSRSTSPCGTSTKS---PNREWEVLQAPAPHLQITNLPPEGVRLPTTRTPPPPLIPSSK 1041
Db 492 LDSGPGSPAKAKASLALPQK---VRPVATQVKTDRKGHSGSESSDSEEEAFAA- 546
QY 1042 TTVASERKPSF--IMGGSISQGP---GTILTSHNQASYQETPKPSVSGISLGLPQOE 1095
Db 547 --SAAQAKPALEKQKASSRKGTTPASATCASTSSHCKAGAV--TSSALLSPALAKGTQRS 603
QY 1096 SAKSATLPYIKQEEFSPRSONSQPEGLLVRAQHEGVVGTAGAIQEGSITRGTTPTSKISV 1155
Db 604 DVDS-----SSESESE---AAPSTPRVQKSG----- 628
QY 1156 ESIPSLRGSITQGTTPALPOTGIPTEALVKG--SISRMPIEDSSPEKGREEAASKGHVIEG 1214
Db 629 -----GKGLQKGAALQGVAVPVHTQKTGPSVKAMAQEDS--ESLEEDSSSE----- 672
QY 1215 KSHILSYDNIKNAREGTRSPRTAHEISLARSYESVEGNKQGMs-----MRESVPVAPLE 1270
Db 673 -----EEDETPAQATPL-----GRLPQAKANPPPTKTPPASA--S 705
QY 1271 GLICRALPRGSHDLKERTVLSGSIQMGTPRATTESFEDGLKYPKQIKRESPPIRAFEG 1330
Db 706 GKAVAAATKGP-----PVPNSIV-SARGORSVPAAGKAGAPATQAKG 748
QY 1331 AITKGKPYDGITTIKEMGRSIEHPIRQDILQTESRKTPVEVVQSTRPIIE----- 1379
Db 749 PVA-GTGEDSESSKEESDEEETPAQ---IKPVGKTSQVRAASAPAKESPCKGAHPCTP 804
QY 1380 ---GSISQGTPIKFDNNSGSAIKHNKVSILITGPKSLSRGMPLEIVPENIKVVERGY 1435
Db 805 GKTGSSATQAPQGTEDSDSSEESDSDTEM---PSAQAIKSPVPVSNRNSSPVAPAPTP 861
QY 1436 EDVKAGETVRSRHTSVVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMSRG 1495
Db 862 EGVQAVNTTKASGTTAQSS-----SSEEDGEDLIPATQPSYALRTSVTTPAALSRA 916
QY 1496 SPMMNRSTDVTIPPKNKTNIHERKSTLTPTQRESIPAKSPVPGVDVYVSHSPFDPHHRGST 1555
Db 917 ASOPSKSEQSRMPKGGKAKAAAQAQTSASAVETLPMMP-----QSAPIQ-----KA 964
QY 1556 AGEVYVWHLR--TOLDPAMP-----FHRALDPAAAAA 1595
Db 965 TNKLGKSKLPEKQOLAPGYPRAPRSSDSDTSSEDEDAKRPOMPKSAHR-LDPDPS-- 1021
QY 1586 LFRQ---LSPTPGYPSQYQLYAMENTQTIINDYIT-----SQOMQVNLRPDVARGLS 1636
Db 1022 --QKETVVEETPTESSEDEWA---PSQSLLSGYMTFGLIVANSQAASKATPRD----- 1070
QY 1637 PREOPLGLPYATPRTGIIIDLNMPTIILVPHPGGTSTPPMDRITVIGTQITFPPRPVNSA 1696
Db 1071 --SNSLASSAPATKNDPDGQKSKS---QHAADTALPKTKKEASSGSTPQPKKKLKKST 1125
QY 1697 SMSPGHPHTLAAAAASARERERERERERERERERERERERERERERERERERERERER 1756
Db 1126 SSSPA-PTQILPNSITQRLLEQ-----AWPLSEAQVQA 1157
```



```

QY 1757 PSPSVRTOETMLQORPSVFOGTNGTSVITPLDPTAQLRIM--PLPAGGPSISQGLPASRY 1814
    | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 1158 SVVKVLTE--LLEQ-----ERLKATEAKESGKSKQKRLSGDLEAGAP----- 1199
QY 1815 NTAADALAALYDAASAPOMDVSKTESKHEAARLEENLRSRAAVSEQ--QLEQKYLE 1872
    | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 1200 -----KNKKKEQPV-----PRASAVSPEKAPMTSKAKSK 1229
QY 1873 VEKRSVQCLYTSSAPPSG---KQPHSSSVYSEAGKDKGPPPKSRYEELRTRGKTTITA 1929
    | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 1230 LDKGSAGKGKSPGPOGAKPKDGLLGKLESQEODPKSKKKKSLAKK----- 1282
QY 1930 ANFIDVITROIADSKDARE-----RGSOSDSSSLSSHRHYETPSDA 1972
    | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : :
Db 1283 -----KDKEKKEKKGKSLAKDSASPIQKKKKKKKKSA 1316

```

Search completed: September 8, 2001, 14:44:04  
Job time: 426 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:34:18 ; Search time 34.32 Seconds  
(without alignments)  
4310.097 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSGGPPNQAFSTEQSRY.....EREPAPLLSAQYETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601:\*

- 1: /SIDSL1/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSL1/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSL1/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSL1/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSL1/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSL1/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSL1/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSL1/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSL1/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSL1/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSL1/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSL1/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSL1/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSL1/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSL1/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSL1/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSL1/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSL1/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSL1/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSL1/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSL1/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDSL1/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	11873.5	93.9	2343	21 AAB12453	Human HNRCR protei
2	11536.5	91.2	2453	21 AAB12454	HNRCR protein sequ
3	4168	33.0	2518	21 AAB40574	Human ORFX ORF338
4	2885.5	22.8	619	17 AAR99738	Retinoid X recepto
5	1879	14.9	1495	18 AAW18226	Transcriptional co
6	520	4.1	3266	21 AAB42491	Human ORFX ORF2255
7	507.5	4.0	2819	22 AAB35408	Human 07CG27 gene
8	404	3.2	3256	21 AAV50976	Human cell cycle p
9	370	2.9	73	15 AAR53557	Thyroid hormone re
10	370	2.9	73	20 AAY40596	Partial amino acid
11	370	2.9	73	20 AAW92403	Human TR-interacti

12	370	2.9	73	20	AAW85115	Thyroid hormone re
13	368	2.9	2843	13	AAR26052	APC gene product 1
14	368	2.9	2843	18	AAW35392	Human adenomatous
15	368	2.9	2843	19	AAW38370	Human adenomatous
16	368	2.9	2860	15	AAAR63507	Adenomatous polypo
17	366	2.9	2843	15	AAAR58634	Adenomatous polypo
18	365	2.9	2742	21	AAAR23012	Human APC protein
19	365	2.9	2842	15	AAAR63508	Adenomatous polypo
20	365	2.9	2843	19	AAW76140	Human APC protein
21	365	2.9	2843	19	AAW76144	Human APC protein
22	365	2.9	2843	21	AAAB23011	Human APC protein
23	365	2.9	2973	19	AAW76821	Human APC protein
24	365	2.9	2973	21	AAW70304	Protein used in ca
25	365	2.9	2973	22	AAW72782	Transcriptional ac
26	364	2.9	2843	16	AAW11922	Adenomatous polypo
27	356.5	2.8	2781	21	AAW57453	Human transcriptio
28	349.5	2.8	3118	22	AAW50362	Human SRCAP. Homo
29	341.5	2.7	2907	21	AAW57452	Human transcriptio
30	329	2.6	2972	22	AAW50363	Human SRCAP. Homo
31	327	2.6	2783	13	AAAR23963	AFP-1 (Ala 2460 Va
32	327	2.6	2971	21	AAAB1231	Human ORFX ORF995
33	326	2.6	2783	13	AAAR23962	AFP-1. Homo sapie
34	325	2.6	1596	18	AAW31347	Rat tumour suppres
35	323.5	2.6	1780	19	AAW53863	Human gravin polyp
36	323.5	2.6	1780	21	AAAB15380	Human gravin prote
37	317.5	2.5	1341	21	AAW85657	Human Acinus L pro
38	316	2.5	3910	14	AAAR38470	ALL-1 protein. Ho
39	316	2.5	3910	16	AAAR66462	ALL-1 (acute lymph
40	313.5	2.5	1719	17	AAAR92100	Human Riz allele D
41	311	2.5	1719	21	AAAB12114	Human mutant Riz p
42	309.5	2.4	2441	16	AAAR79054	CREB binding prote
43	309.5	2.4	2441	21	AAAY94252	Mouse nuclear CREB
44	309	2.4	1719	21	AAAB12029	Human RB-interacti
45	308.5	2.4	2441	19	AAW40058	Cellular transcript

## ALIGNMENTS

RESULT 1

AA12453

ID AAB12453 standard; Protein; 2343 AA.

AC AAB12453;

DT 24-OCT-2000 (first entry)

XX

DE Human HNRCR protein SEQ ID NO:20.

XX

XX Human; HNRCR; nuclear receptor coreceptor.

OS Homo sapiens.

XX

PN CN1250094-A.

XX

XX 12-APR-2000.

XX

XX 06-OCT-1998; 98CN-0120919.

XX

XX 06-OCT-1998; 98CN-0120919.

XX

XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.

XX

XX Yu L, Tu Q, Zhao Y;

XX

XX WPI; 2000-400830/35.

XX

XX N-PSDB; AAA60629.

XX

XX Preparation of new human keron acceptor co-repressor coding series and

XX

XX the polypeptide -

XX

XX Claim 4; Page 25-30; 58pp; Chinese.

XX



CC The present sequence represents a human homologue of nuclear receptor  
CC coreceptor (HNRCR).

XX	SQ	Sequence	2343 AA;
----	----	----------	----------

Query Match 93.9%; Score 11873.5; DB 21; Length 2343;  
Best Local Similarity 94.3%; Pred. NO. 0;  
Matches 2316; Conservative 5; Mismatches 6; Indels 129; Gaps 5;

<b>Qy</b>	<b>1</b>	MSSGYPNQGASTEGSRYPHVSQVTFNTRHQGEFAVPDYRSSHLEVSQSALLQQQ	<b>60</b>
		:	
<b>Db</b>	<b>1</b>	msssvypnqdgafstegsrphvskvtfpsthhqq-----	<b>36</b>

Qy 61 QQQQLRRRPSLLSEFHGSDRPPQERTSYEFHFGSPVDHDSLESKRPRLEQVSDSHQ 120

Dp 37 ----- 36

Qy 121 RVSAANVLPLVHPLPEGLRASADAKKDPAFGGKHEAPSSPISGQPCGDDONASPSKLSKEE 180  
 |||||  
 27

```

Qy 181 LIQMDRVDREIAKVEQQILKLKKQQOLEEAAKPPPEKPPPEQKHSIVQIIY 240
      |||||

```

Qy	241	DENRKAEEAHKIFEGGLGPKVELPLYNQPSDTKVYHENIKTNQVMRKKLLIFFKRRNHAR	3000
----	-----	--	------

DB	132	denrkkakeanklreglgpksvelplnyndpsatckvynenikctnqvmrkklllrrkkfnnar	19
Qy	301	KREQKICQYDQLMEAWKKVDRIENPPRRKAKESKTREYYEKQFPEIKRQEQERFQ	360

Db 192 kreqkicarydqlmeawekkvdiennprkakesktreyyekqfpeirkqreqqerfq 251

Qy 361 RVGORGAGLSAT1ARSEHEISEIIDGLSEQENNEKQMRQLSVIPPMFDFAEQRRVKFINM 420

Db	252	rvqgrgagisatiaeheiselidgiseqesnekqmqslsvppmmfdaeqrkvafinn	311
QY	421	NGLMEDPMKVYKDRQFMNVWTDHEKEIEFKDKFIQHPKNFGLIASYLERKSVDPDCVLYYYL	480

	312	ngImedpmkvykdraqmnnvwtdehekeIfdkkfIqhpnfgllasylerksvpdcvlyyyL	371
Db			
Oy	481	TKKNENYKALVRRNYGKRGRNNOOIARPSOEKKVEEKEEDKAETKEKPEEKDEEEKDE	540

Db	372	tkknykalvrnygrrgnqgiarpsqeeekdae	431
O:	541	wpdcknkmkncvctpcmtg	500

Db 432 kedskentekkdigdtaeeteereqatprgrktansgrrkgritrsmtneaaaaaa 491

492 aaateepppplppppp1stspvetsrwtteeneevakg1vehgrnwaaiakmvgtksea 551

Qy	661	QCkNFYFNyKRrHnIdNlLlQhKqKtSRkPReeRdVSQcSVASvtsAQeDEIEASNEE	720
		=====	
Db	552	qckNFyfnYkrRrHnIdNlLlqhkKqKtSRkPReeRdVSqCSvtsvtsAQeDEIEASNEE	611

Qy	721	ENP	EDSE-----	-----	VEAVKPS	EDSP	ENAT	SRGNT	EP	AVELE	PTT	ET	AP	ST	764						
Db	612	enp	edse	gaenss	dt	es	ap	sp	pv	avkps	edsp	enat	sr	gnt	ep	avele	ptt	et	ap	st	671

Qy	765	PSLAVPSTKPAEDES	VETQVND	SI	SAETA	EQMD	VQQE	HS	AE	EGSV	CDPP	PPAT	KADSV	DV	824
Db	672	pslavpstkpaedes	vetqvnds	is	aetaeqm	dvqqe	hsae	egsv	cdpp	patkadsv	dv	731			

QY	825	EVRYPENHASKVEGNTKERDLDRASBKVEPRDEDLVVAQIINARPEPOSDNDSATCS	884
Dh	712	EVRYPENHASKVEGNTKERDLDRASBKVEPRDEDLVVAQIINARPEPOSDNDSATCS	791



QY 2025 SPQOQLPSSQAEGQGVPRTHRLITLADHICQITODFARNQVSSOTPOQPTSTFQNS 2084  
 Db 1929 spqqlpssqaegmgvprthrilitladhicqitqdfarnqvssqtpqgqptstfqn 1988  
 QY 2085 PSALVSTPVRTKTSNRYSPEQAQSVHHQRPGRSVSPENLYDKSRGSPGKSPERSHVSS 2144  
 Db 1989 psalvstpvrtktsnryspesqasvhhqrpgrsvspenlydkrgsrpgkspershvss 2048  
 QY 2145 EPEYPIPPQPVVHEKQDSLLLSQRAEPAEQRNDAKSPGSIYILPSFPTKLENTSPM 2204  
 Db 2049 epeyipppqvvhkqdslllsqraepaegrndarspgsisylpsfftklentspm 2108  
 QY 2205 VKSKQEIFRKLNSGGGSDMAAQAQCTEIFNLPAVTTSGSVSRGHSFADPASNIGLE 2264  
 Db 2109 vkskqelfrklnsgggdsmaaaqctefnlpavttsgsvsrghsfadpsnigle 2168  
 QY 2265 DIIRKALMGSDDDKVEDHGVVMSOPMGVPGTANTSVVTSGETRREGDPSPHSGGVCKP 2324  
 Db 2169 diirkalmgstddkvedhgvmsopmgvpgtantsvvtsgetrreedpsphs-gvckp 2227  
 QY 2325 KLISKSNSRKSPIPGQYLGTERPSVSVHSEGYHROTQGWAWEDRPSSTGSGTQFP 2384  
 Db 2228 klisknsrkskipggylgterpsvsvhsegyhrotqgwawedrpsstgsgtqfp 2287  
 QY 2385 YNPITMRMLSTPTPIACAPSVAQNQAAPHQONRIWEREPAPLLSAQYETLSDSD 2440  
 Db 2288 ynpitmrmlstptpiacapsavnqaaphqonriwerepapllsaqyetlsdssd 2343

## RESULT 2

AA12454  
 ID AAB12454 standard; Protein; 2453 AA.

AC AAB12454;

DT 24-OCT-2000 (first entry)

XX HNRCR protein sequence.

XX Human; HNRCR; nuclear receptor coreceptor.

XX Unidentified.

XX CN1250094-A.

XX 12-APR-2000.

XX 06-OCT-1998; 98CN-0120919.

XX 06-OCT-1998; 98CN-0120919.

XX (XINH-) XINHANGPU FUDAN GENE ENG CO LTD SHANGHA.

XX Yu L, Tu Q, Zhao Y;

XX WPI; 2000-400830/35.

XX N-PSDB; AAA60630.

XX Preparation of new human kernon acceptor co-repressor coding series and the polypeptide -

XX Example 2; Fig 2; 58pp; Chinese.

XX The present invention describes a human homologue of nuclear receptor coreceptor (HNRCR). The present sequence represents an HNRCR protein sequence used in comparison with the human HNRCR.

XX Sequence 2453 AA;

Query Match 91.2%; Score 11536.5; DB 21; Length 2453;  
 Best Local Similarity 91.0%; Pred. No. 0;  
 Matches 2245; Conservative 79; Mismatches 101; Indels 43; Gaps 12;

QY 1 MSSGYPPNOCAGFSTEQSRYPHVSQVTEPNTRHQEQFAVDPDYRSSHLEVSQAQLQQQ 60  
 Db 1 mssgyppnqgafstegrypshvsqvtfnarhqefavpdyrsshlevsqaqlqqq 60  
 QY 61 QQQQLRRRPSLLSEFFHFGSDRPERRTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQ 120  
 Db 61 qqqlrrrpsllsefhpgrsdprerrsgyefhpgspvdhdsleskrprleqvdsdshf 120  
 QY 121 RVSAVLPLVHPLPEGILRASADAKDPAGCKHEAPSSPISGQPCGDDONASPKLSKEE 180  
 Db 121 rvsavlpvlhplpegilrsaanakdpagvckheapspslsgpcgddqnaspsklskee 180  
 QY 181 LIQMDRVDRREIAKVEQOILKLLKKQOOLEEAAKPEPEKPPVPPVPEQHRSTVQIYI 240  
 Db 181 liqmdrvdrreiakevqilklkkqqleeeakppekpvppvpeqhrstviyiy 240  
 QY 241 DENKKAEEAHKIPGEGIPKVELPLYNQSDTKVYHNIKTQVMRKLLILFFRRNHAR 300  
 Db 241 denrkkaeeahkifeglpkvelplynqpsdkvyniktqvmrkllilffrrnhar 300  
 QY 301 KOREQKICORYDOLMEAEKWKVDRIENNPRAKESKTREYVEKOFPEIRKOREQOERFQ 360  
 Db 301 kreqkicorydqlmeawekvdrinennprkakestreyyekofpeirkoreqerfq 360  
 QY 361 RVGQAGLGSATIAARSEHEISEIIDGLSEONNEKQMRQLSVIPPMFADQRVVKFINM 420  
 Db 361 rvqraglsatiarseheiseiiddglseqennekqmrqlsvipmmfdaeqrrvkvfinm 420  
 QY 421 NGLMEDPMKVKYKQFMVNTDHEKEIFKDFIQHPKNFGLIASYLERKSPVDCVLYYL 480  
 Db 421 nglmedpmkvykqfmmvntdhekeifkdfiqhpkngfliasylerksvpdcvlyyl 480  
 QY 481 TKNENYKALVRNRYGKRGNNQIARPSQEEKVEEKEEDKAEKTEKKEEKDEEKDE 540  
 Db 481 tknenykalvrnrygkrgnngqiarpsqeekeveeekedkaektekeekdkdeekdd 540  
 QY 541 KEDSKENTKEKDKIDGTAETETEEREQATPRGKTANSQGRKGRITRSMTEAAAAA 600  
 Db 541 kedskentkekdki dgtateteteereqatprgkktansqgrkgvtrtsmtaataa-aa 599  
 QY 601 AAATEPEPPPLPPPPPEPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP 660  
 Db 600 aaatepeppplpppppepistpvtsttsttsttsttsttsttsttsttsttsttsttst 659  
 QY 661 QCKNFYNYKRRHNLNLQOHKQKTSRKPREERDVQCESVASTVSAQDEDEDLEASNEE 720  
 Db 660 qcknfynykrhnlndllqhkqksrkpreervdvsqcesvastvsaqededleasnee 719  
 QY 721 ENPEDSE-----VEAVKPSDESDPENATSRGNTPEPAVELEPTTAPSTS 764  
 Db 720 enpedsegaenssdtesapspsveaakssedssenaasrgntpepvaeleatdpapcas 779  
 QY 765 PSALVSTKPAEDSVETQVNDISIAETAEQMDVQDQHEAEGSVCDPPPAPKADSDV 824  
 Db 780 psavpttkpaeresveaqvtdsasaetaepmdvdheecgaegsvldppaptkadvdp 839  
 QY 825 EVRYPENHASKVEDNTKERDLDRASEKVEPRDEDLVVAQOINAQRPPEQSDNDSSATCS 884  
 Db 840 emqvptaskgegd-akerdlestektardedvvvaegi--erpepqsdssatcs 896  
 QY 885 ADEDVDCGEPRQRMPMDSKPDLNPTGSLVSSPLKPNPLDLPLQOHRAAVIPPVYSCT 944  
 Db 897 adevgdgeprqrvfmdakpslltpggsllisspiknplldlpqlqhraavipmvsct 956  
 QY 945 PCNIPITGVSGYALYQRIKAMHESALLEQRQORQEQIDLECRSSTSPGCTSKSPNREW 1004  
 Db 957 pcnipitgvsgyalqyrihikamhesalleqrqrqeqvdlccrscstspcstskspnrew 1016  
 QY 1005 EVLQAPAPHQLITNLPEGVRLPTTRTPPPPLIPSSKTTTVASEKPSFIMGSIQGTPT 1064  
 Db 1017 evlqapahqvltnlpegvrlpttrtprrppplipssktttvasekpsfimgsgisqgtpt 1076



QY 1065 YLTSHNQASYTQETPKPSVSGSISLGLPQROESAKSATLPIYIKQEEFSPRSQNSQEGILV 1124  
Db 1077 yllshnqa-yqeaqpkvsgsisglprqgestkaapityikqeefspqnsqegilv 1135  
QY 1125 RAQHEGVVGTAGAIQBSITRGNPTSKISVESIPSRGSIQTQTPALPQGTGIPTEALVK 1184  
Db 1136 raqhegvvgtagavqegsitrgtpaaskisvetissirgsitqgtpalpqaqiptealvk 1195  
QY 1185 GSISRMPTEDSSPKGREAAKHVIEYEGSKGSHLSVDNFKNAREGTRSPRTAHEISLK 1244  
Db 1196 gvisrmpieesspekvreeaasknviyegskgnihlydniknaregtrsrptanemsik 1255  
QY 1245 RSYVESGNIKQGMRESPPVAPLEGLICRALPRGSPHSDLKERTVLGSIQGTPTAT 1304  
Db 1256 rsyavegsikqgmresppvapeleglicralprgshsdlkertvlgsgimgtprat 1315  
QY 1305 TESFEDGLKYPKQIKRSPPTIRAFEGAITKGPYDGTITIKEMGRSHEIPRQDILQES 1364  
Db 1316 aesfedgkypkqikresppirafegaikgkpydgtittikemgrsheiprqdiltqes 1375  
QY 1365 RKTPEVQOSTRIIEGTSIQGTPIKFDNNSGOSAIKHNVKSLITGPKSLGRGMPPLEIVP 1424  
Db 1376 rktpevqgstriiegsisqgtpkfdnmsgsaikhnvkslitgpkslgrgm--leivp 1433  
QY 1425 ENIKVVERGKYEDYKAGETVRSRHTSVVSGPSVLSRTLHEAPKAQLSPGIYDTSARRT 1484  
Db 1434 enikvvergkyedvkagetrarhtsvvsgpsvlsrtlheapkaqlspgilyddssarrt 1493  
QY 1485 PVSQNTMSRSGPMNRTSDVTIPNKSHTNHERKSTLTPTQRESIPAKSPVPGVDVPSVH 1544  
Db 1494 pvsqntisrsgpmnrtsdvs--ssksasheksstltptqresipakspvpgvdpvsh 1551  
QY 1545 SPFPDHRHSGPAGVYVSHLPTQLDPAMPFHRALDPAALYLFQRLSPTPGYPQSQQLY 1604  
Db 1552 spfdphrssaagevyrshlptdampfhraldp-aaayllqrlsptgtpgypsqqly 1610  
QY 1605 AMENTRQILNDYITSQOMQVNLRPDVARGLSPREQIPLGYPATRGIIIDLTNNPPTILV 1664  
Db 1611 amentrqilndyitseqmqvnlrpdvtrglspreqiplgypatrgiildltnnpptilv 1670  
QY 1665 PHPGGTSTPPMDRTYIPGTOITPPRPYNSASMSGPHPTLAAAASAEEREREREK-- 1722  
Db 1671 phpggtstppmdrtypgtvttpprpynaaslsphgnptlhaaaaeererereker 1730  
QY 1723 -----ERERERETAAASSDLRPSQEQGRPCSGHYVRSPSPVTRTQBTMLQQRPSV 1774  
Db 1731 erererereriiaaapadlylrpsqeqgrpgshgyvrspsvtrtqbtliqqrpsv 1790  
QY 1775 FQGTNGTSVITPLDPTQRLRIMPLPAGGPSISQGLPASRYNTAADALVDAASAPQM 1834  
Db 1791 fqgtngtsvitpldptqqlrmlpipsgppsissgglpasryntaadalaalvdaasapqm 1850  
QY 1835 DVSKTKESKHEAARLEENLRSRAVSEQQOOLEKTEVEKRSVQCLYTSSAFPSGKQP 1894  
Db 1851 dvsktkeskeaarleennlrsraavseqqglekntevekrsvqcltssafpsgkqp 1910  
QY 1895 HSSVYVSAGDKGPPPKRSYEELIRGKTTITAAFNIDVIIITRQIASOKDAREGSSQS 1954  
Db 1911 hasvvyseagdkgpppkrsyeelirgkttitaanfivdiitrqiasdkdaregssqs 1970  
QY 1955 SDSSSSLSHRYETPSDAIEVISPASSAPPQEKLOQYQPVVVKANQAEENDPTROYEGPL 2014  
Db 1971 sdsasslshryetasdaievispassappqekpqqayqpdmvkandaeendptroyegpl 2030  
QY 2015 HHYRQOESPSPOQ--LPPSSQAEQMGVPRTHRLTLADHICQIITQDFARNQVSSQT 2072  
Db 2031 hhyrqqespspqoq--lppssqaeqmgvprthrlltladhicqiiitqdfarnqv---- 2086  
QY 2073 PQQPTFTFQNSALVSTPVTRTKTSNRYSPEQASVHHQRPGRSVSPENLVKRSGRS 2132  
Db 2087 psqastftfqsalsstpvtrtktsrnysspesqsvlhprrpvrspenlvkrsgrs 2146  
QY 2133 PGKSPERSHVSSEPYEPISSPQVPVHVKQDSLILLSSQGAEPAEQRNDRKSPGSIYLP 2192

Db 2147 pgkspershipsepyepispqpqpavhekqdsmlllsqrgvdpaeqrdsrsrpsgsisylp 2206  
QY 2193 SFFTKLENTSPWVSKKQOEIFERKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSRGH 2252  
Db 2207 sfftklestpmvkskkqeifrklinsgggdsmaaaqpgteifnlpavttsgavsrsh 2266  
QY 2253 SFADPASNLGLEDIIRKALMGSDDKVEDHGVMQPMGVVPGTANTSVVTSQGTREEG 2312  
Db 2267 sfadpasnlglediirkalmgssfdkvedhgvmshpvgimpssastsvtssaardeg 2326  
QY 2313 DPSPHSGGVCKPKLISKSNSRKSPIPGQGYLGTERPSSVSVHSEGDYHROPFGWAVE 2372  
Db 2327 epspha-gvckpklinknsrkskspipggsylgterpsvsvhsegdyhrqtpgwawe 2385  
QY 2373 DRPSSTGSTOPPYNPLTMRMLSSSTPTPIACAPSAVNAQAAPHQONRIWEREPAPLLSAQY 2432  
Db 2386 drpsstgstgstopnpltrmlsstptqiacaapsaitqaaphqonriwerrepapllsaqy 2445  
QY 2433 ETLSDDSD 2440  
Db 2446 etlsdsdd 2453  
RESULT 3  
AAB40574  
ID AAB40574 standard; Protein; 2518 AA.  
XX AAB40574;  
AC AAB40574;  
XX 08-FEB-2001 (first entry)  
XX Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;  
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;  
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
XX cholesterol ester storage; systemic lupus erythematosus; infection;  
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;  
XX thrombosis; contraceptive.  
XX Homo sapiens.  
XX WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
XX 02-APR-1999; 99US-0127636.  
XX 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0340763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC74783.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
XX useful for treating e.g. cancers, proliferative disorders,  
XX neurodegenerative disorders and cardiovascular disease -



Claim 11; Page 784-790; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnerary; anticiporatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antidiabetic; hypotensive; dermatological; immunosuppressive; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 2518 AA;

Query Match 33.08; Score 4168; DB 21; Length 2518;

Best Local Similarity 40.88; Pred. No. 1.4e-219;

Matches 1101; Conservative 340; Mismatches 784; Indels 476; Gaps 108;

QY 17 QSRYPHSHVQVTFPTNRHQEFAPVDYRSSHLEVSQASQLQQOQQOQLRRRPSILSEFH 76  
DB 16 epyrphshslsypvdiarhtdvgileq-hhsrdyashlspgslqpprrrlsef 73  
QY 77 PGSDRQPB--RTSYEPFHPGSPVDHDSLESKRPRLEQVSDSHFQRVAALVPLHPLP 134  
DB 74 pgnrsqelhlrpeshsylpelgkxsefieskrprlelpld-----pllrpsp 122  
QY 135 ---EGLRA-SADAKDPAFGKKEAPSSPTSGQPCGDDONASPSKLSKEELIQSDMRDVR 190  
DB 123 llatqpagseelldkdrsltgkle-pvsppphtdpelelvprlskeelilqmdrvdr 181  
QY 191 EIAKVEQOILKLLKKQOOLEEAAKPEPEKPVSPVPEQKHSRIVQIYIDENRKAAPEA 250  
DB 182 eitmvqgiskllkkqgleeaaakpepekpvspppieskhrslvqliydenrkkaea 241  
QY 251 HKIFEGLPKVELPLYNQPSDTKYHENIKTNQVWRKLLIFFRRNRHARKQROKTCQR 310  
DB 242 hrilegipqvelplynqpsdtrqyheniklnqamrkllilyfrrnrharkqwekfcoqr 301  
QY 311 YDOLMEAWKVDRIENNRKAKESKTRYEKQFPEIRKQEQOERFO-RVGORGAGL 369  
DB 302 ydqlmeawekvrienrrrakesvreyekqfpeirkqelqermqgrvggrsgl 361  
QY 370 SATIARSEHETSEIDGLSEQENNEKQRLSVIPPMFQAEQRRVFINNGLMEQPMK 429  
DB 362 smsaarsehevseilglseqenlekqmrqlavipmlydadqgrikfinnmgimadpmk 421  
QY 430 VYKDRQPMVWTDHEIFDKDFTQHPKNFGLIASYLERKSVPCVLYYYITKKNENYKA 489  
DB 422 vykdrqpmvnmwsekektfrefmqhpkngfliasflerktvaecvlyyitkknnyks 481  
QY 490 LVRRNYGKRRGNQIARPSQEEVEKEED---KAEKTEKKEEKDEEKDEKESKE 546  
DB 482 lvrrsy-rrrgsqgqqqqqqqqqqqqqqpmprsqeekdekekeakeeekpeve 540  
QY 547 NTKE---KDKIDGTA-BETEEREQATPRGRKTANSQGRKGRITRSMTNEAAAAASAAAA 602  
DB 541 ndkedlllkektddsgednkeavaagkrktansqgrkgritrsmneanseaitpq 600  
QY 603 ATEEPPLPPPLPPPISTEPVETSRWTEEMEVAKGLVEHGRNWAIAIAKVMGKTSBAQC 662  
DB 583 drkitstpre-----iaksphstvpbehphipisvhehlirgsvdvlyrshiplafdpctsi 1638

DB 601 qs-----aelasmelnessrteeemetakkgllhgrnwsaiarmvgsktvsqc 650  
QY 663 KNFYFNKRRNRNLNLQOHKOKTSRKPRERDVSQCESVASTVSA-----QEDEDIEAS- 717  
DB 651 knfyfnykkrqnldeilqghklmekernarrkkkapaaseeaaafppvvdeemeasg 710  
QY 718 ---NEEENPEDE-----VEAVK----- 732  
DB 711 vsgneemveeaealhasgnevprgecsgpatvnnssdtesipshteaakdtgngkpkp 770  
QY 733 -----PSEDSPENATSRGNTPEPAVELE---PTTETAPSTSPSLAVPSTKPADE 778  
DB 771 patigadgppgpprtprtsrapieptpaseatgaptppap-psps-appvvvypkcek 828  
QY 779 SVETQVNDISIAETAEQMDVQDQSHSAEESVCDPPPPATK--ADSDVDEVRVPENHASKV 836  
DB 829 eeetaaappv-----eegeekppaaaelavdtgkaeepvksecteaa 871  
QY 837 EGDNTKERDLDR-----SEKVEPRDELVVAQOINQAQRPPEPQSDNDNSATCSADE 887  
DB 872 eegpakgkdaaeataegaalkaekkgsgrrattaksga---pq-dsdsatcsade 926  
QY 888 --DVDGEPERMPMDSKPSLLNPTGSLV-SSPLKPNPLDLPLQLQHRAAVIPPMWVSC 944  
DB 927 vdeaeeggdknrlsp---rpsalltptgdpranaqpk--pidllkqlkqraaaiippi----- 977  
QY 945 PCNIPIGTPVSGYALYQRIKAMHESALLE-----EORQROEQIDLECRSSTS 992  
DB 978 -----qvtkvheppredaaptkppappppppqnlqpesdapqpgss 1019  
QY 993 PCGTSKSP-----NRE-----W-----EVLQAPAP-----QLITNL 1018  
DB 1020 prgksrppadpkaeataeaaklpgdpccwtsglpgfvpvprevikasphadpsafsa 1079  
QY 1019 PEGVRLP-----TTRPRPPPLPLPSSKTTVASBK-PSFI--MGSSISQGTGTYLTSHN 1070  
DB 1080 ppgheplglghdrtarpvlprprrtlnppplissakhpavlerqigalsqg-----msvql 1134  
QY 1071 QASVTQETPKPSVSGISILGLPRQOESAKSATLPVYKOEFPSPSONSOPEGLLV-RAQHE 1129  
DB 1135 hvpysehakap-vpvtmgllpmdpkklapfsgvkceqlsprgagppeslgvptaea 1193  
QY 1130 GVVRGTA-GAIOEGSITRGTTFTSKISVESIPSLRGSITQGTTPALPQTGIPTEALVKGSGIS 1188  
DB 1194 svlrgtalgsyvggsitkgipstrvpsdsaityrgsichgtpa-----dvlykgkit 1245  
QY 1189 RMPEDSSP--EKGREEAASGHVIEYEGSKSHILSYDNKNA-----REGTRSPRTAHEI 1241  
DB 1246 rilgedspsrldrgedsplkghviyegkghvlsyeggmavtqcskedgrssagpphet 1305  
QY 1242 SL-KRSVESVEGNIKQGMRESVPVSAPLEGLICRALP--RGSPHSDLKERTVLSGSTMQ 1298  
DB 1306 aaprttydmegrvgrais-----sasieglmraipprhshp-hlkeqghirgsitq 1358  
QY 1299 GTPRATTESEFEDGL-KYPKQIKRES-----PPIRAFEGAI-----TKGRP-YDG-ITT 1343  
DB 1359 giprsyveaqedyirreakllkregtppppps-dltaeyktqalglklkpaheglvat 1418  
QY 1344 IKEMGRSITHEIPRODILTQESRKTPVVOSTRPIEGSISOGTPIKFDNN-SGQSAIKHN 1402  
DB 1419 vkeagrsiheipreel-----rhtpelaprlksgitgdtgtpkydtagstgskhd 1473  
QY 1403 VKSLITGPSKLSRCMPPLEIIVPENIKVVVERKGYEDVKAGETVRSRHTSVWSSGSPVLRS 1462  
DB 1474 vrsliagsgrtfppvhpldvmad-aralaracye-----eslkrrpgtasssggsiarga 1527  
QY 1463 ---LHEAPKAQLSPGIYDDTSARRTPVSQNTMGRSGSPMMNRTSDVTIP-----PNKSPNH 1515  
DB 1528 pviyelpgkpkqspltyedhga-----pfaghlprgsptvtreptprlqegslssskasq 1582  
QY 1516 ERKSTLPTQRESIPAKSPVPGVDVPSH--SPDPHHRGSTAGEVYVSHLPTOLDP-AM 1572  
DB 1583 drkitstpre-----iaksphstvpbehphipisvhehlirgsvdvlyrshiplafdpctsi 1638















Db 794 kdvdeyrrslvhevqkppqdvddssppskkrr-----mhvdfdictkreynrssrqi 848  
Qy 206 QOOLBEEAAKPP-----EPEKPV-SP-----PPVEQK-----HRsIVQII 239  
Db 849 sedsertgssvrhgshfededpigsrllsvkspkvdskvdpysnitvreesalkfnp 908  
Qy 240 YDENRKK--AEEAHKIFGLGPKVDELPLYNOPSDFKVVHENIK-----TNOVMRK 287  
Db 909 ydsrrreqmadma-----kiklslvnsedelnrwdsqmkqdagrdvdfpsnlikr 959  
Qy 288 KILILFFKRNHARKQREOKICORYDOLMEAKKVDRIENPRRKAKESKREYVEKQFP 347  
Db 960 dsi-----kfrsvrdlepvepsds-----edgch-----shprasalysessrflldr 1008  
Qy 348 EIRKOREQOERFORVGORGA-----GLSATIARSEHEISEIDGL--SQEN-----392  
Db 1009 e-dklirderlssalernkfyfaldktitpdkallierakslssreenwsfildwsr 1067  
Qy 393 -----NEKOMRQLSVP---PMFPAEORRVKFINNGLMEDPMKVVYKDRFMNVWTOHE 444  
Db 1068 fanfrnnkdkekvdapripipswy-mkkkkir--tdsegkmdkdkedhkeee-----qer 1119  
Qy 445 KEIFKDKFIQHPKFNGLIASYLERKSPDCVLYLYLTTKKNYKALVRNKGKRRGRNQ 504  
Db 1120 qifasfih-----ssifeq-----dskrlqlherkeedsdfisrlygkqtsega 1167  
Qy 505 IARPSQEEKVEEKEDKAETKEEKKDEEKDEKEDSKENTKEKIDGTAEETER 564  
Db 1168 sttdsiqevpvlfshrfmelt-rmqkkkkkdkqpkvekqedenhpktpesapenks 1226  
Qy 565 EQATPRG-----RKYANSQGRKKGRITRSMWNEAASAAAAAATEE 606  
Db 1227 elktppsppsvvtvlesapsalektgtktveaplvteektvapatvseeakpasep 1286  
Qy 607 pppp-----Lpp-----613  
Db 1287 apapveqleqvdlppgadpkeaaampagveegsgdppylldakptpgasfsqaesnv 1346  
Qy 614 PPEPTSTPEVETSRWTEEMEYAK-----KGLVEHGRNAAIA-----651  
Db 1347 dpepdstqlspkpadkseeanepkaekpadataadaepdanqkaeapesqppasedlevdp 1406  
Qy 652 -----KMGVTKSAQCKNFYFNKRRHNDLNLQOHHKQTSRKPRED-----v 696  
Db 1407 pvaakdkkpnkskrtppvgaavsvivekpvtrkseridreklrksnspgreaqklllek 1466  
Qy 697 SQCESVASTSAQEDEDIE-----715  
Db 1467 meakitrtasknaadliehpelplstrrrrvrsyvatmgdhenrsrvkpevqprv 1526  
Qy 716 -----ASNEBENPEDSEVEAVKPSD--SPENATS 743  
Db 1527 trklrelqeaaavtprgrbpktrrradeeeneakepaetlkppegvrsprskt 1586  
Qy 744 -----RGNTPEAVEL---EPTTETAPTSTPSLAVPSTKPADESVEVQVNDISIAE 791  
Db 1587 aagggpggkkgnekpvdatrpeattevgpglgvkesmepkaaeaeasegqdrkdag 1646  
Qy 792 TAEQMDVQDQEHSAEAGSVCCPPATKADSDVDEVVRVPENHASKVEGDNTKER-----844  
Db 1647 tdkn-----ppetapvevvekkpapekkskrgrsrnsrlavdksa 1688  
Qy 845 DLDRASEKVEPRD-----EDLVAAQOINAQRPE-PQSDNDSATCSADE-DVDGPE 894  
Db 1689 slknvdaavprgaaagageresgvva--vspekseapkedgissqklsdvpdpdkpe 1746  
Qy 895 RORPMDKPSL-----LNPTGSILVSSP--LKPNPLDLPOLQ 931  
Db 1747 kedvsasopspeatqlakmeleqavehiaklaeasaaaykadapeglapedrdkphq 1806  
Qy 932 ----HRAAVIPPM---VSCPTCNPIPIGTPVSGYALYQRIKAMHESALLEEORQOEID 984  
Db 1807 asetelaaagsiindisgepenfapppypyg-----esqtd 1843

Qy 985 LECRSSTSPCGTSKSPNREWEVLQAPAHQIITNPEGVRLPT---TRTPRP--PPLIP 1038  
Db 1844 lq-----ppaga-----qalqpseegmetdeavsgileteaessrrpvpnapd--p 1888  
Qy 1039 SKTTVASEKPSFIMGSIQSQTPTGYLTSHN--QASVTOETPKPSVSGISIGLPRQOES 1096  
Db 1889 sagptdtke-----argnsse-----tshsvpeakgsk-----vevllvrkdkg 1928  
Qy 1097 AKSATLPIYKO-----EEFSPRSONSQOPELLVRAQHEGVVVRGTAGAIQE---GS 1143  
Db 1929 rqtctsrkrntnkvvavpveshvpesnaqges---paanegtvtvqheapeqgeqse 1985  
Qy 1144 ITRGTP-----TSKIVESIPSLRGSITQCTPALPOTGPTALVRKGSISRMPIEDSSSPK 1199  
Db 1986 kphstppgsctsdls--kipstens-sqeisveert-ptkasvpdpdlp--pppapavd 2038  
Qy 1200 GREEAASGHVYIEG-----KSGHILSYDNINKNAREGTRSPRTAHEISLKRSYESVEGNI 1254  
Db 2039 eepqarfvrhsliesdpvtppsdpisipitlpsvtaaklspva---sggiphqspptkv 2095  
Qy 1255 KOGMSMRSP--VSAPLBGLICRALPRGSPHSDLKE---RTVLSSGSIQGTTPRATTES 1307  
Db 2096 tewitrcpgepraqscp-----spalppdtkasdvdsstlrlkilmdpkyvsatsvts 2150  
Qy 1308 FEDGLKYPKQIK---RESPPIRAFEGAITKGRPYDGITTIKEMGRSIIHEIPRQDILTOES 1364  
Db 2151 vttaiaepysaapclheapp-----ppvdskkpleekctappvttnns--eigasevlvaad 2203  
Qy 1365 RKTPTVWOSTRPIIEGSIQSQTPIKFD--NNSQSAIKNNKSLITGPGSKLSCMGMPLEIV 1423  
Db 2204 ke--kvapviapkitsvisr--mpvsiidlensqkiklappatqtlglvsaltglvnsvl 2260  
Qy 1424 PEN-LKVVERGKYEDVAGETVRSRHTSVWS--SGP-SVLRSTLHEAPKAQLSPGIYDDT 1479  
Db 2261 pvnalkgpykgsvtlik-----slvtpagpvnvllkgpvnv-----ltgvpv 2301  
Qy 1480 SARRTPV-----SYQNTMSRSGSPMMNRT--SDVTIPPNKSTNHERKSTLTP-----1523  
Db 2302 nvlttpevnatvgtnaagptvnaaasavnatasavtvtagavtaasggvtagtvtgvmag 2361  
Qy 1524 -----TQRESIPAKSPV-PGVDPVVSHPFPPHHRGSTA-----GEVWSHL- 1564  
Db 2362 aviapstkckgrasanensrfhpgsmpvldrdpad--agsgaglrvtseggvllvsyg 2418  
Qy 1565 -----PTOLDPAMPFHRALDPAAYL-FOROLS-----PTGYPSPQYQLY 1604  
Db 2419 kkteqprisaki---sqippasamdiefgqsvksqvkpdsvtasqppskg-papagay 2474  
Qy 1605 AMENTROTILNDYITTSQOMQVNLDPDVARGLSPR---EQP-----LGLPYPATRG-----1652  
Db 2475 anvathstlv---ltaq--tynaspvisvskadrpslekpepihlsvstptvggtkvvl 2529  
Qy 1653 IDLTNMPPT-----ILVPHPGGISTPPMDRITYITPGTITFPPRPYN-SASMSGHPH 1705  
Db 2530 tggintppvlvhnqilvtlpsivttkkldapvtliketkvl---qpanlgstltphhpa 2586  
Qy 1706 LAAASASEREREREK---ERERERIAAASDYLVRPGSEOPGRPGSHGVYVRSPSPVR 1762  
Db 2587 lpsklptevnhvpsgspipadrtvshlaaakldah-----sprpspgpspsfrashps-s 2641  
Qy 1763 TQETMLOORPSVF---QGTNGTSVITPDLPTAQLRIMPLPAGGSPISQGLPAS-----R 1813  
Db 2642 tastalstnatvmlaaglpvqpfissihp-eqsvlmp---phsitqlvslhlsqgevr 2696  
Qy 1814 YNTAADALAALVDAASAPQMDVSKTESKHEAARLEENLRSSAAVSEQOOLEOKTLEV 1873  
Db 2697 mnt-----ptlpsitysirpealh-----spraplpqpg-----i 2726  
Qy 1874 EKRSVQOCLYTSAPSGKPPQPHSSVYSEAGKDKGPPPKSRYEELRTRGTTTAAFI 1933  
Db 2727 evrapqrastpqapagpv-----alasqhp-----eeevhhlvvaratapvq 2771















Db	2758	gedgkalkesakqtcpaasvt--gsr--rrprapresaqaedlaqfkdpaghtees	2814
Qy	2165	LLLLSORGAPAPQRNDARSPGISYLPSPFFTKLENTS-----PMVSKSKQEIFRKLNS	2218
Db	2815	m-----tddkttkipckss-----peledatsskrprtrraqkvevkeella	2857
Qy	2219	SG-----GGSDMAAQAQ-----GTEIFNLPA-----VTVSGSVSSRGHSFADPASNGLGED	2265
Db	2858	vgklqtsgettthdkpevggkgtkafkqapkrnvdaedvlgstrprapkekaqpled	2917
Qy	2266	IIRKALMGSPDDKVEDHGVVMQPMGVVPGTAN----TSVVTSGETRRREGGPPSPHSGVC	2322
Db	2918	-----lasfqe-----lsqtgghteelangaadftsaqkqtpdsgkplksrrvl	2963
Qy	2323	K-PKLISKSNRSKSKSPIPQCYLQTERP	2350
Db	2964	rapkvepvqdvvstrdpvksgaksnstslp	2992

RESULT 9  
AARS3557  
ID AARS3557 standard; protein; 73 AA.  
XX  
XX AARS3557;  
XX  
XX  
DT 10-FEB-1995 (first entry)  
XX  
XX Thyroid hormone receptor-interacting protein.  
XX  
XX  
XX nuclear thyroid hormone interacting proteins; TR; JLI; JLI2;  
KW transcriptional coactivator; treatment; diagnosis;  
KW thyroid related disorders; modulation; thyroid hormone receptor;  
KW nuclear hormone receptor; isolation.

44	Chimeric Homo sapiens.
OS	Chimeric Bacterial sp.
OS	Chimeric Bacterial sp.

XX XX  
PN PN  
WO9410338-A.

XX PD 11-MAY-1994.

XX  
PF 29-OCT-1993; 93WO-US10443.

XX 30-OCT-1992; 92US-0969136.  
PR

PA (GEHO ) GEN HOSPITAL CORP.

FA (GENCO) GEN HOSPITAL CORP.  
XX  
XX  
PI Lee JW, Moore DD;

FI  
XX  
Lee JW, MOORE DD,  
WPI: 1994-199808/24.

DR  
DR WFI; 1994-199806/24.  
DR N-PSDB; AAQ63705.  
XX  
XX

aa PT PT Nuclear hormone receptor interacting polypeptides, esp. thyroid hormone-interacting proteins (TRs) - for identifying proteins

PT useful in treatment and diagnosis of thyroid related disorders b

XX  
PS Claim 35; Page 61; 105pp; English.

xx  
cc This sequence shows the partial amino acid sequence of a thyroid

CC hormone (TR) interacting protein. TR-interacting proteins physically associate with thyroid hormone receptor. Nearly all the fusion c

CC showed very strong dependence on hormone activation. The protein can be used in an *in vivo* trap system for the isolation of proteins that are secreted by the same cell type. The protein can be used in an *in vivo* trap system for the isolation of proteins that are secreted by the same cell type.

CC which associate with any nuclear hormone receptor. The proteins  
CC Abs may be used to treat or diagnose thyroid disorders; and to  
CC modulate thyroid hormone receptor activity.

CC	modulate thyroid hormone receptor activity.
XX	
SC	sequence 72 aa.

**SQ** sequence / 3 AA;

Query Match 2.9%; Score 370; DB 15; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84



Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSPADPASNLGLIEDIIRKALMGSDDKVEDHGVVMSQPM 2290  
|||||  
Db 1 pgtelfnlpavttsgvssrghsfadpasnlglediiirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
|||||  
Db 61 gvpvgtantssvvt 73

RESULT 10  
AAY40596  
ID AAY40596 standard; protein; 73 AA.  
XX  
AC AAY40596;  
XX  
DT 03-DEC-1999 (first entry)  
XX  
DE Partial amino acid sequence of TR-interacting protein S243b.  
XX  
KW JLI protein; thyroid hormone receptor; trap assay; therapeutic;  
KW nuclear hormone receptor protein; TR-interacting protein; human;  
KW thyroid disorder.  
XX  
OS Homo sapiens.  
XX  
PN US5962256-A.  
XX  
PD 05-OCT-1999.  
XX  
PF 06-JUN-1995; 95US-0471613.  
XX  
PR 04-APR-1994; 94US-0222719.  
PR 30-OCT-1992; 92US-0969136.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
XX  
PI Lee JW, Moore DD;  
XX  
DR WPI; 1999-571268/48.  
DR N-PSDB; AAZ07582.  
XX  
PT Purified DNA comprising a sequence encoding a protein which  
PT specifically interacts with a thyroid hormone receptor -  
XX  
PS Disclosure; Fig 28; 68pp; English.  
XX  
CC The invention provides a JLI protein which specifically interacts with a  
CC thyroid hormone receptor in an in vivo trap assay. The JLI protein can  
CC be recombinantly produced by standard recombinant methodology. The JLI  
CC protein is used in a trap assay for determining whether a test protein  
CC is capable of interacting with a nuclear hormone receptor protein, and  
CC may also be used as a therapeutic peptide for treating thyroid  
CC disorders. Sequences AAY40572-596 represent partial amino acid sequences  
CC of thyroid hormone receptor (TR)-interacting proteins.  
XX  
SQ Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSPADPASNLGLIEDIIRKALMGSDDKVEDHGVVMSQPM 2290  
|||||  
Db 1 pgtelfnlpavttsgvssrghsfadpasnlglediiirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
|||||  
Db 61 gvpvgtantssvvt 73

RESULT 12  
AAW85115  
ID AAW85115 standard; protein; 73 AA.  
XX  
AC AAW85115;  
XX  
DT 09-FEB-1999 (first entry)  
XX  
DE Thyroid hormone receptor-interacting protein S243b partial sequence.  
XX

RESULT 11  
AAW92403  
ID AAW92403 standard; protein; 73 AA.  
XX  
AC AAW92403;  
XX  
DT 21-APR-1999 (first entry)  
XX  
DE Human TR-interacting protein S243b.  
XX  
KW Thyroid hormone receptor-interacting protein; TR-interacting protein;  
KW JLI; human; interaction trap assay; treatment; thyroid disorder; S243b;  
KW hyperthyroidism.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Protein 1..73  
FT /note= "Partial coding sequence, no start or stop  
FT codons given"  
XX  
PN US5866686-A.  
XX  
PD 02-FEB-1999.  
XX  
PF 06-JUN-1995; 95US-0470925.  
XX  
PR 04-APR-1994; 94US-0222719.  
PR 30-OCT-1992; 92US-0969136.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
XX  
PI Lee JW, Moore DD;  
XX  
DR WPI; 1999-142013/12.  
DR N-PSDB; AAX01894.  
XX  
PT JLI protein - that interacts with thyroid hormone receptor  
XX  
PS Disclosure; Fig 28; 67pp; English.  
XX  
CC This sequence represents a human thyroid hormone receptor (TR-interacting  
CC protein) which is found to interact with a novel human JLI protein in an  
CC in vivo interaction trap assay. The JLI protein is potentially useful for  
CC the treatment of hyperthyroidism or thyroid disorders.  
XX  
SQ Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEIFNLPAVTTSGVSSRGHSPADPASNLGLIEDIIRKALMGSDDKVEDHGVVMSQPM 2290  
|||||  
Db 1 pgtelfnlpavttsgvssrghsfadpasnlglediiirkalmgsfddkvedhgvvmsqpm 60  
|||||

QY 2291 GVVPGTANTSVVT 2303  
|||||  
Db 61 gvpvgtantssvvt 73

RESULT 12  
AAW85115  
ID AAW85115 standard; protein; 73 AA.  
XX  
AC AAW85115;  
XX  
DT 09-FEB-1999 (first entry)  
XX  
DE Thyroid hormone receptor-interacting protein S243b partial sequence.  
XX



KW Thyroid hormone receptor-interacting protein; S243b;  
 KW nuclear hormone receptor protein; screen.  
 OS Homo sapiens.

XX US5846711-A.  
 XX  
 XX 08-DEC-1998.

XX 04-APR-1994; 94US-0222719.  
 XX  
 XX 04-APR-1994; 94US-0222719.  
 PR 30-OCT-1992; 92US-0969136.

XX (GEHO ) GEN HOSPITAL CORP.  
 XX PA  
 XX PI Lee JW, Moore DD;

XX WPI; 1999-059040/05.  
 DR N-PSDB; AAV82605.  
 XX

XX Screening assay for nuclear hormone receptor modulators - using  
 XX cells containing reporter gene construct

XX Disclosure; Fig 28; 69pp; English.

XX The present sequence represents thyroid hormone receptor-interacting  
 CC protein S243b. The protein was identified using the method of the  
 CC invention. The method is used to determine if a test protein is capable  
 CC of interacting with a nuclear hormone receptor protein in a  
 CC ligand-dependent manner. The method is used especially to screen for  
 CC proteins that interact with thyroid hormone receptors in a  
 CC ligand-dependent or ligand-sensitive manner.

XX Sequence 73 AA;

Query Match 2.9%; Score 370; DB 20; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2231 PGTEFNLPATVTSVSSRGHSFADPASNLGLIEDIIRKALMGSDFKVEDHGVVMSQPM 2290  
 Db 1 pgtelnlpavttsvssrghsfadpasnlgledliirkalmgsfdkvedhgvvmsqpm 60

QY 2291 GVPVGTANTSVVT 2303

Db 61 gvpvgtantsvvt 73

RESULT 13  
 AAR26052

ID AAR26052 standard; Protein; 2843 AA.

XX AAR26052;

XX 28-JAN-1993 (first entry)

XX APC gene product in familial adenomatous polyposis.

XX neoplasm; cancer; oncogene; tumour; growth; detection; diagnosis;  
 KW prognosis; treatment; sporadic colorectal carcinomas; ss.

XX Homo sapiens.

XX WO9213103-A.

XX 06-AUG-1992.

XX 16-JAN-1992; 92WO-US00376.

XX 16-JAN-1991; 91GB-0000963.

PR 08-AUG-1991; 91US-0741940.

XX (CANC-) CANCER INST.  
 PA (ICIL ) IMPERIAL CHEM IND PLC.  
 PA (UJO ) UNIV JOHNS HOPKINS.  
 PA (UTAH ) UNIV UTAH.

XX Albertsen H, Anand R, Carlson ML, Groden JL, Hedge PJ;  
 PI Joslyng, Kinzler KW, Markham A, Nakamura Y, Thliveris A;  
 PI Vogelstein B, Whitehl, Markham AF;

XX WPI; 1992-284685/34.  
 DR N-PSDB; AAQ27234.

XX Detection of somatic and germ-line alterations of human APC gene  
 PT - used to diagnose, treat and study familial adenomatous  
 PT polyposis and sporadic colorectal cancer

XX Disclosure; Page 47; 132pp; English.

XX This sequence is encoded by the APC (Adenomatous Polyposis Coli)  
 CC gene associated with tumorigenesis, found on chromosome 5q.  
 CC The sequence may be mutated by deletions insertions, inversions, or  
 CC point mutations of the gene. The APC gene is expressed in most normal  
 CC tissues as well suggesting that APC is a tumour suppressor.

XX Sequence 2843 AA;

Query Match 2.9%; Score 368; DB 13; Length 2843;  
 Best Local Similarity 19.4%; Pred. No. 1.8e-11;  
 Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

QY 347 PEIRKQREOE-RFORVGORGAGLSATIARSEH-----EISEIIDGLSEQENNEK 395  
 Db 981 psiesyddedskfcygyqypadlahkhsanhdndgdeldtptnlskysdedsgr 1040

QY 396 QMRQLSVIPPMFMDAQRVKFINNGLMEDPMKYKDRQFMVMTDHE--KEIFKDFEI 453  
 Db 1041 q-----spsqnerwarpkhiedelkqseqgrngstttypvtestdkhl 1087

QY 454 QHPKNFGLIASYLERKSVPCVLYLTKNENKALVRRNYKRRGNQOQIARPSOEK 513  
 Db 1088 kfqphfg-----qqecvspy-----rsrgangsetnrvsnhngnqvsglqce- 1132

QY 514 VEEKEEKAETK--BEEKDEEKEDESKENTKEKIDGTABETEEREQATPRG 571  
 Db 1133 -dyeddktynseryseeqheeeerptnysikyn-eekrhvdpidyslk yatdipss 1190

QY 572 R-----KTANSQGRKGRITRSMNEAAS-----AAAAATEEPPPLPPP 614  
 Db 1191 qkqsfkskssgsktshmsststpsnakrnqlhpsaqsrgsqgqkaatck 1250

QY 615 PEPISTEPVEV-----SRWT-----EEMEVAKKGLVEHGRNWAATKM--- 653  
 Db 1251 vssinqetigtvcvedtpicfscsslsssaedeigcndtqeadsantlqlaekgk 1310

QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLQHK-----QKTSR 688  
 Db 1311 igtrsaedpvsevpavsqhprtkssrlqgsslsesarhkavfsgakspksaqtpk 1370

QY 689 KPRE-----RDVSCBSVASTVSAQDEDEIASNEENPEDSEVEAVKPE--DSPEN 740  
 Db 1371 sppehyvqetplmfrctsv-ssidsfesrsiasvqpcsgmvsiglisldipdsppg 1429

QY 741 A--TSRGNTPEAVELEPTTETAPSTPSLAVPTKPADESVETQVNDISIAREQMDV 798  
 Db 1430 tmppsrsktppp---ppqatqkrevpknkaptakresgpkqaavnaavg-rvqvlpda 1485

QY 799 DQQHSAAE-----GSCV-----DPPPATKADSVDEVR---VPEN-HASKVEGDN 840  
 Db 1486 dtllhfatestpdpfscsslsalsldpfqk----dvelrmpvqendngneteseq 1541

QY 841 TKE--RDLDRASEKVEPRDEDLVVAQINAQRPEFPQSDSSATCSADEVD----- 890







Example 1; Columns 33-52; 140pp; English.

The present sequence is the human adenomatous Polyposis coli (APC) gene product, which was used in the development of a novel method of diagnosing or prognosing an APC gene associated neoplastic tissue. The method comprises comparing APC gene coding sequences on mRNA in a tumour tissue, to APC gene coding sequences or mRNA in a non-neoplastic tissue, where a difference indicates an APC gene associated neoplasia of the tumour tissue.

APC is a tumour repressor expressed in most normal tissues. APC mutations are found in familial adenomatous polyposis and sporadic colorectal cancer patients. The method enables mutations to be detected to provide an indication of predisposition to cancer.

AA	Sequence	2843 AA;
SQ		

Query Match	2.9%;	Score 368;	DB 18;	Length 2843;
Best Local Similarity	19.4%;	Pred. No. 1.8e-11;		
Matches 418;	Conservative 281;	Mismatches 797;	Indels 664;	Gaps
QY	347	PEIRKQEQE-RFORVGORGAGLSATARSEH-----EISEIIDLGLSQQENNEK	395	
DB	981	psiesydesdeskfcsygydpadlahkhsanhdmdngdeltdpinyiskysdeqlnsgr	1040	
QY	396	QMRQLSVIPPMFDABQRVRKFINNGLMDPMKYKQRFMNVWTDHE--KEIFKDKFI	453	
DB	1041	q-----spsqnerwarphkhiiedeikqsegrsqnqsttvpvtestddkhl	1087	
QY	454	QHPKNEGLIASYLERKSPDCLVYLYLLKKNENYKALVRYNYKRGNGNOQIARPSOEK	513	
DB	1088	kfaphfg-----qqecvspy-----rsrgangsetnrvgshnqinqvsgslcqe-	1132	
QY	514	VESEKDEKAEKTEK--EEEEKDEEKEDEKESKENTKEKDIDGTABETEERQQAIPRG	571	
DB	1133	-dveddkptnyseyseceqheeeerptnysikyn-cekhrvdqpidyslkvatidpss	1190	
QY	572	R-----KYANSQGRKGRITRSTWTEMAAAS-----AAAAATEPEPPPPPPP	614	
DB	1191	qkgsfksksgqssktehmssssentstpsnknqrnlhpsaqsrgqpkaatk	1250	
QY	615	PEPISIEPVET-----SRWT-----EEMEVAKKGLVEHGRNWAATAKM---	653	
DB	1251	vssingetiqtqvcedpicfrcsslsasaeideicnqtqeadantlqlaiekkg	1310	
QY	654	VGTKSE-----AOCKNFYFYKRRHNDNLQOHK-----OQTSR	688	
DB	1311	igrtsaedpvsevpavsqhprtksrlgsslsesasrkhavefpsgakspksqgaqtpk	1370	
QY	689	KPREE-----RDVSCESVASTVSAQDEDEIEASNESENPEDESEVAVKPESE--DSPEN	740	
DB	1371	sppehyvqetplmfrctsv-slsdfesriassvgqpcsgmvgslispedlpdpgpgq	1429	
QY	741	A--TSRGNTPEPAVELEPTTETAPSTPSLAVPSTKPADESVEVQVNDISIAETAQMDV	798	
DB	1430	tmppsrsktppp--ppqtaqtktrevpknkaptakresgpkqaavnaavq-rvqvlpda	1485	
QY	799	DQEHGAEE-----GSVC-----DPPATKADSDVDVEVR-----VPEN-HASKVGEDN	840	
DB	1486	dtllhfatestpdgfcscsslsalsldpfiqk----dveirlmpvqndngneteseq	1541	
QY	841	TKG--RDLDRASEKVEPRDELVAQQINAQRPQSDNDSSATCSADEVDV-----	890	
DB	1542	pkesnqenqekektldsekdl-----ddsd-----ddlleileecelis	1581	
QY	891	-----GEPERQRMF-----PMDSKPSLLNPTGSILVSPSLKPNPLDLPLQGHRAAVI	937	
DB	1582	amptksrkqkqpaktasklpppvarkpsql-pvykllpsq---nrl---qpqkhsyft	1633	
QY	938	P-----PWWSC---TPCNTPIGTGPVSGYAL-----YORHIKAMHE	969	
DB	1634	pgdmprrvycveqtpinfstalsldltiesppnelaagevrgqagsgefktdtipte	1693	

Qy	970	SAL----	LEBORQREQOIDL	ECRSSTS	PCGTSKSPNREVL-----	1007	
Db	1694	grstdeagqktsv	tipeclddnkaeegil	laecinsampk	gkshkprfvrkkmldqvqqa	1753	
Qy	1008	--QAPQHOLITNL	PEGVRLTPTRPR	PPPLI-----	-----p	1038	
Db	1754	sassapnk---	hqldgkkktpsvk	pipqnteytrvr	knadsknlnaervfsdnkd	1810	
Qy	1039	SKKTVTASEKPSF	-----IMGSGIS	-----QCTPGVYLTS	SHNOASTYQETP	1079	
Db	1811	skknlknnskdfnd	klnnedrvrgsfad	ghyhtpiegtp--	ycfsrnd-----	1860	
Qy	1080	KPSVGSISLGLP	QROESAKSATPLY	LKQEFSPRSON	SOPEGLLVRAHQEGVVRGTAGAI	1139	
Db	1861	--sissldfdddv	lsrekaelrkaken	kesekavtshte--	ltsndqsa---nktqai	1913	
Qy	1140	QEGSITRGTPTS	KISVESPSLRGSI	TQCTPALPOTGI	PTREALV-----G	1185	
Db	1914	akqipnrgqkpi	lqkqs-----	tfpqskdipdrgaat	deklnqfaientpvcfshns	1967	
Qy	1186	SISRM-----	PIEDSSPEKGREEA	SKGHVIEGKSGHI	-----	1219	
Db	1968	slssldidgenn	kenepiketepdsg	epskpqasgyapks	fhvedtpvcfsrnssl	2027	
Qy	1220	--LSYDN-----	TKNAREGTRSPRT	AHETLSLRKSVES	VEGNTKQCSMRBSPVSAPL	1269	
Db	2028	sslsldseddl	lqecissampkkk	psrllkgndkeshpr	nmggilgedtldldkdrpd	2087	
Qy	1270	EGLICRALPRGS	PHSDLKERTVL	SGIMQCTPRATTE	SPEDGLKYPKOLIKRESPIRAFE	1329	
Db	2088	se---hqlspdsen	fwkaiqegansiv	ssllqaaaaac---	lsrqasdsdsilsk	2139	
Qy	1330	GAITKGRPYDGI	TIKEMGRSHEIP	RODILTQESKRTP	VEVQSTRPIIEGSIQGTPIK	1389	
Db	2140	sgislgpf-----	hltpdqeekpftsn	kngprilk-----	pgekstletkk	2180	
Qy	1390	FDNNS-QGSAIK	HNVKSLTGP---	SKLSRGM-PPELIV	PENIKVVRGKYEDVKAGET	1443	
Db	2181	ieseskglggk	vykylsitgvrns	eslsgmqkqplq---	anmpsirsq-----	2227	
Qy	1444	VRSHTSVWSG	PSVLKSTLHEAP	KALQSPGIYD	TTSARRTPVSQNTMSRGSPPMNR	1503	
Db	2228	-----rtmih-----	lpgv-----	lpgv-----	lpgv-----	2251	
Qy	1504	DVTIPPKNSTN	HHERKSTLTP-	QRESIPAK-SPVPG	VDPVVSHSPDPHHRGSGAGEYVW	1561	
Db	2252	plktpakspseg	qtattspgakpsv	kseiselpvarqts	qigsskapsgrsgr-----	2305	
Qy	1562	SHLTQDLPAMP	HRALDPAANA	YLQROLSPTPG	YPSQYLIYAMENTROTILNDYITSQ	1621	
Db	2306	-----	-----	-----	-----	2319	
Qy	1622	QMQVNLRPDVA	--RGLSPREQL	GLPYPATRGI	IDLTMNPTILVPHPGTSTPPMDRI	1678	
Db	2320	plqspgrnslsp	grnglslppnk-----	lslqprt-ssp	staskssgsgkm-----	2364	
Qy	1679	TYIPGTQITPP	PPYPYNSAS	PGHPHAAAAA	SARERERERE-----	1726	
Db	2365	sy-----	tspgqrmeqnlit	--kqglsknass	iprsesaekglnqmnngankkvel	2416	
Qy	1727	ERIAAASDYL	LRPGSEQPRG	SGHYVRSPSP	SVRTQETMLQORPSVFOGTNCTSVITP	1786	
Db	2417	srmsstks-----	gsesd-----	rseprlvirg	stfikeaps-----	2449	
Qy	1787	LDPTAQIRIMPL	PAGPSPISQ	GLPASRYWTAD	ALAAALVDAASAPQMDVKTSKHEA	1846	
Db	2450	--pflr-rklees	asfesls-----	psprpaspr	strsqatpvlspslpdmsls-thssvqag	2502	
Qy	1847	A--FLENLRAS	NAAVSEQOOL	BQKLTLEYEK	RSVQCLYTSSAFPSPKPPHSSVYSEAG	1904	
Db	2503	gwrklpnl---	sptleyn	dngrprakhdlar	-----shsesprl	pnrsrgtwkrehs	2552
Qy	1905	KDGPPPKRSY	EELTRGKTTIT	TAAN-----	FIDVITRQIASDKDARERGSSSD	1956	



Db 2553 khssalp--rvstwrtsasessseksedkhvnsisgkqskengvsakg 2610  
 QY 1957 SSSSLSHRYE--TPSDALEVISPASSPAPQEKLOTQYQPEVVVKANQANDTRQYEGDLH 2015  
 Db 2611 twrkikeneifpntnsqcvsgatngaesktliyqmapavsk---tedvvrriedcpin 2667  
 QY 2016 HYRPOQESPSQOQLPP-----SSQAEGMGVPRTHRLITLADHI 2055  
 Db 2668 apr-----srsptgntppvidsvsekanpnikdskdnqakqnvngsvpm--rtvglenri 2722  
 QY 2036 COIITQDFARNQVSSQTPOQPT--STFQNGPSALVST---PVRTKTSNRYSPESQAQSV 2110  
 Db 2723 tsfiqvd-----apdqgkteikpgnupvpvsetnespivert-----pfssssss 2768  
 QY 2111 HHQRP-----GSRVSPENLVDSRGSRPKSPERSHVSSEPEPISPPQVPVH--EKQDS 2164  
 Db 2769 khsspsgtvaarvtfnfny-----npsprksasadsarpsqip--tpvnnntkkrrds 2818

RESULT 15  
 AAW38370  
 ID AAW38370 standard; Protein; 2843 AA.  
 AC AAW38370;  
 XX  
 DT 08-APR-1998 (first entry)  
 DE Human adenomatous Polyposis coli gene product.  
 XX  
 KW Human; adenomatous Polyposis coli; APC; diagnosis; prognosis;  
 KW neoplastic tissue; tumour tissue; tumour repressor; mutation;  
 KW sporadic colorectal cancer; detection.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US5691454-A.  
 XX  
 XX 25-NOV-1997.  
 XX  
 XX 25-MAY-1995; 95US-0452654.  
 XX  
 PR 16-JAN-1991; 91GB-0000975.  
 PR 16-JAN-1991; 91GB-0000962.  
 PR 16-JAN-1991; 91GB-0000963.  
 PR 16-JAN-1991; 91GB-0000974.  
 XX  
 PA (CANC-) CANCER INST.  
 PA (ICIL) IMPERIAL CHEM IND PLC.  
 PA (UYJO) UNIV JOHNS HOPKINS.  
 PA (UTAH) UNIV UTAH.  
 XX  
 PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;  
 PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;  
 PI Vogelstein B, White RL;  
 XX  
 DR WPI; 1998-017712/02.  
 DR N-PSDB; AAT96153.  
 XX  
 XX Antibodies to normal and mutant adenomatous polyposis coli proteins  
 PT - useful for detecting genetic predisposition to cancer  
 XX  
 PS Example 1; Columns 33-52; 107pp; English.  
 CC  
 CC The present sequence is the human adenomatous Polyposis coli (APC)  
 CC gene product, which was used in the development of a novel method  
 CC of diagnosing or prognosing an APC gene associated neoplastic  
 CC tissue. The method comprises comparing APC gene coding sequences or  
 CC mRNA in a tumour tissue, to APC gene coding sequences or mRNA in a  
 CC non-neoplastic tissue, where a difference indicates an APC gene  
 CC associated neoplasia of the tumour tissue.  
 CC APC is a tumour repressor expressed in most normal tissues. APC  
 CC mutations are found in familial adenomatous polyposis and sporadic

CC colorectal cancer patients. The method enables mutations to be  
 CC detected to provide an indication of predisposition to cancer.  
 XX  
 SQ Sequence 2843 AA;  
 Query Match 2.9%; Score 368; DB 19; Length 2843;  
 Best Local Similarity 19.4%; Pred. No. 1.8e-11;  
 Matches 418; Conservative 281; Mismatches 737; Indels 664; Gaps 98;  
 QY 347 PEIRKQEQE--RFQVQORGAGLSATIARSEH-----EISEIIDGLSEQENNEK 395  
 Db 981 psiesyddeskkcsygyypadiahkhsaumdndgeidtpinyslkysdeqlnsgr 1040  
 QY 396 QMRQLSVIPPMFQAEQRRVKFINWGLMEDPMKYKDRQFMNVWTDHE--KEIFKDKFI 453  
 Db 1041 q-----spqnerwarpkhiefedkqsegqrndsttppvtestddkhl 1087  
 QY 454 QHPKNFGLIASYLERKSPVDCVLYYLTKKNENYKALVRRNYKRRGNOQIARPSOEK 513  
 Db 1088 kfqbhfg-----qqecvspy-----rsrgangsetnrvgsnhginqnvsqldqe- 1132  
 QY 514 VERKEEDKAETKK--EEKKDEEEDKEDSKENTKEKIDGTAEETEEREOATPRG 571  
 Db 1133 -dyeddktynyseryeedqheeeerptnysikyn-eekrhvdpidysikyatdipss 1190  
 QY 572 R-----KTANSQGRKGRITRSMTEAAAAA-----AAAAATEEPPLPPP 614  
 Db 1191 qkqfsfsssgsgsktehmssssststpsnakrqnlhpsasqsgsqgqaatck 1250  
 QY 615 PEPISTPEPVET-----SRWT-----EEEMEVAKKGLVEHGRNMAAIAKM--- 653  
 Db 1251 vssinqetiqtvyvedtpicfscrsissisaedegcngtqeadsantliaekgk 1310  
 QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLLQHK-----QKTSR 688  
 Db 1311 igtrsaedpvsevpavaghprrtkksrliqgssalsesarthkavefpgsakspksqgatk 1370  
 QY 689 KPREE-----RDVSOQESVASTVSAQEDIEDIASENEENPEDESEAVEAKPSE--DSPEN 740  
 Db 1371 sppehyvqetplmfscrtsv--ssidsfessrsiasvsgepcsmvgsilspdsldpdpq 1429  
 QY 741 A--TSRGTEPAVELEPTTAPSTPSLAVSTKPADESVETQVNDISIAETAQMDV 798  
 Db 1430 tmppsrsktppp---ppqtaqtkevknkaptakresgpkqaavnaavq-rvqvlpda 1485  
 QY 799 DQEQHSAEE---CSVC-----DPPPATKADSDVDVEVR---VPEN-HASKVEGDN 840  
 Db 1486 dtllhfatestpdgfsccsslsalsldepfikq----dveirlmpvpqengdngneteseq 1541  
 QY 841 TKE--RDLDRASKEVPRDELDVVAQIQAORPEPQPSDNDSSATCSADEVD----- 890  
 Db 1542 pkesnenqeakektidsekdl-----ddsd-----dddieileecils 1581  
 QY 891 -----GEPERQRMF-----PMDSKPSLNLPTGSLVSSPLKPNPLDLPLQAHRAVI 937  
 Db 1582 amptksrkgkpkpaqtasklpppvarkpsql-pvylklpsq-----nrl---qpqkhvst 1633  
 QY 938 P-----PMVSC---TPCNIPITGTPVSGYAL-----YQRHKAMHE 969  
 Db 1634 pgddmrvyvcgtplnfstatsltsditiesppnelaagevrgvgaqsegekrdtipte 1693  
 QY 970 SAL-----LEEQRQREQIDLEKRSSTPCGTSKSPNREVEL----- 1007  
 Db 1694 grstdeaggtktsvtipeiddnkaeeegdilaecinsampkgshkpfvrkkinndqvqqa 1753  
 QY 1008 ---OPAPHQLITNLPQGVRLPTTRTPRPPLI-----p 1038  
 Db 1754 sassaapnk---nqldgkklkktspkvpiponteytrvrknadsknnlnaervfsdnk 1810  
 QY 1039 SSKTTVASEKPSF-----IMGSGIS-----QGTPTGTYLTHNQASVYQETTP 1079  
 Db 1811 skkqlknnskdndklnpnedrvrgsfafdsphhytplegtp--ycfsrnd----- 1860



Qy 1080 KPSVGSISGLPQESAKSATLPYIKQEEFSPRQNSQEGCLLVRAQHEGVVRGTAGAI 1139  
Db 1861 --slsldfddvdlrkaelrkakenekeseakvtshte--ltsnqgsa--nktqai 1913  
Qy 1140 QEGSTTRGTPYKSVESIPSLRSGITQGTALPQGTGIPTEALVK-----G 1185  
Db 1914 akqplnrgqpkpiliqkgs-----tfqsskdlpdrgaatdekliqfaientpvcfshns 1967  
Qy 1186 SISRM-----PIEDSPEKREAAKSGHVIYEGKSGHI----- 1219  
Db 1968 slsldidqennkenepiketepdsggepskpaasyapksfhvedtpvcfsrnsel 2027  
Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLKRYSYESVEGNIKQGMRESPPVSAPL 1269  
Db 2028 ssldsdldllqecissampkkkpsrlkgdnekhspnmngilgedltldkdiqrpd 2087  
Qy 1270 EGLICRALPRGSPHDLKERVLSGSIQGTTPRATTSFEDGLKYPQIKRESPIRAFE 1329  
Db 2088 se---hglspdsenfkwalkqegansivsslhqaaaac-----lsrqassdsilslk 2139  
Qy 1330 CAITKGRPYDGTITIKEMGRSIIHPIRQDILTQESRKTPEVVQSTRPIEGSISQGTPIK 1389  
Db 2140 sgislgapf-----hltpdeekpftcnskqprlik-----pgekstletkk 2180  
Qy 1390 FDNNS-GOSAIKHNKSLITGP-----SKLSRGM-PPLEIVPENIKVVERGKYEDVKAGET 1443  
Db 2181 lseskgikggkkykslitgkvrnsesigqmkpqlq---anmpsirg----- 2227  
Qy 1444 VRSRHTSVSSGSPVSLSTLHEAPKAQLSPGIYDDTSARRTPVSQNTMRGSPMMNRTS 1503  
Db 2228 -----rtmih-----ipgv-----rnsstspvskkkgp 2251  
Qy 1504 DVTIPPNNKSTNHERKSLTPT-QRESIPAK-SPVPDGVDPVVSHPSPDPHHRGSTAGEVYW 1561  
Db 2252 plktpakspsegatatsprgakpskelspvarqtqigsskapsrgsr----- 2305  
Qy 1562 SHLPTQLDPAMPFRHALDPAAAYLFORQLSPTEGYPYSQVQLYAMENTROTILNDYITSQ 1621  
Db 2306 -----dstpsrpaqpl----- 2319  
Qy 1622 QMOYNLRPDVA---RGLSPREOPGLPYPATRGIDITNMPTTLVPHPGGTSTPPMDRI 1678  
Db 2320 plqspgrnsispgngslspnkn-----lsqprt-ssbstastkssgsgkm 2364  
Qy 1679 TIPTGTQITPPRPNYSASMGHPTHLAAAASAEERERERE-----KERER 1726  
Db 2365 sy-----tspgrmqsqnit--kqtglknassiprsesaskglngmnngngankkvel 2416  
Qy 1727 ERIAAASDLYLRPGSEQPGRPGSHGVRSFSPSVRTQETMLQORPSVFQGTNGTSTVTP 1786  
Db 2417 srmsstkss-----gsesd-----rserpvlvrqstfikeaps----- 2449  
Qy 1787 LDPTAQLRIMPLPAGPSISQGLPASRYNTAADLALVDAASAPQMDVSKTESKHEA 1846  
Db 2450 --ptlr-rkleesasfsls---psrpsaptrsqadtqvlslpaldmsls-thssvqag 2502  
Qy 1847 A-RLEENLRSRAAASVSEQOLEQKTELEVKRSVQCLYTSSAPFSGKPPQHSSVYSEAG 1904  
Db 2503 gwrklpnl---sptieyngdrpakrhdia-----shsesprlpnhrsgtwkrehs 2552  
Qy 1905 KDGPPPKPSRYEELRTRGKTTITTAAN-----FIDVITITRQIASDKDARERGSSSD 1956  
Db 2553 khssslp--rvstwrtrgssslsassesekaksekdekhvnsisgtkqskengvsaakg 2610  
Qy 1957 SSSLSLSHRYE-TPSDAIEIVSPASSAPQEKLTQVQPEVVKANQENPTTRYEGPLH 2015  
Db 2611 twrkikeneftsntstqvtssgatngaesaktliymapavsk---tedvwriedcpin 2667  
Qy 2016 HYRPOQESPSQQLPP-----SSQAEQMGQVPRTHRLITLADHI 2055  
Db 2668 npr-----sgrsptgntppvidsvsekanpnikdkdnqakngvngsvpm---rtvglenrl 2722

Qy 2056 CQIIQTDFARNQVSSQTPQOPPT--STFQNSPSALVST---PVRTKTSNRYSPESQAOVS 2110  
Db 2723 tsfiqvd-----apdqgtktpqnnpvpvsetnespivert-----pfssssss 2768  
Qy 2111 HHQRP-----GSRVSPENLVDSRGRSPKSPERSHVSSEPISEPPOVPPVHH--EKODS 2164  
Db 2769 khsspsgtvaarvtpfny-----npsprksadstsrpsqip-tpvnnntkkrd 2818  
RESULT 16  
AAR63507  
ID AAR63507 standard; Protein; 2860 AA.  
XX  
XX AAR63507;  
XX  
DT 30-MAY-1995 (first entry)  
XX  
DE Adenomatous polyposis coli tumour repressor.  
XX  
KW Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;  
KW familial adenomatous polyposis; cancer diagnosis and prognosis;  
KW tumorigenesis suppression.  
XX  
OS Homo sapiens.  
XX  
XX US5352775-A.  
XX  
PN 04-OCT-1994.  
PD  
XX  
PF 08-AUG-1991; 91US-0741940.  
XX  
XX 16-JAN-1991; 91GB-0000962.  
PR 16-JAN-1991; 91GB-0000963.  
PR 16-JAN-1991; 91GB-0000974.  
PR 16-JAN-1991; 91GB-0000975.  
XX  
PA (CANC-) CANCER INST.  
PA (ICIL ) IMPERIAL CHEM IND.  
PA (UJO ) UNIV. JOHNS HOPKINS.  
PA (UTAH ) UNIV UTAH.  
XX  
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;  
PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;  
PI Vogelstein B, White RL;  
XX  
XX WPI; 1994-316233/39.  
DR N-PSDB; AAO72297.  
XX  
PT New human adenomatous polyposis coli DNA encoding tumour  
PT repressor - and derived primers and probes for diagnosis,  
PT prognosis and treatment of cancer  
XX  
PS Claim 3; Fig 7; 113pp; English.  
XX  
CC AAO72297 is a cDNA isolated from the human adenomatous polyposis  
CC coli (APC) gene, it encodes the tumour repressors described in AAR63507  
CC and AAR63508. Determination of alterations in APC or its expression  
CC products, can be used for the diagnosis and prognosis of cancer e.g.  
CC colorectal, lung and breast tumours; and for determining predisposition  
CC to certain cancers such as familial adenomatous polyposis (FAP) and  
CC Gardner's syndrome. The wild type APC gene (or a part of it) can be  
CC used therapeutically to restore gene function, while primers and probes  
CC derived from the cDNA (AAO72333-400 and AAO72541-568) can be used to  
CC detect mutations. Also APC proteins or analogues can be administered to  
CC compensate for a defective gene, and epithelial cells, or transgenic  
CC animals carrying a mutated APC allele are useful for detecting  
CC therapeutic agents able to suppress tumorigenesis.  
XX  
SQ Sequence 2860 AA;  
Query Match 2.9%; Score 368; DB 15; Length 2860;  
Best Local Similarity 19.4%; Pred. No. 1.9e-11;



Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

QY	347	PETRKOREQE	-RFQVRGORGAGLSATTARSEH-----ELSEITDGLSQEENNEK	395
QY	348	QY	QY	QY
QY	349	QY	QY	QY
QY	350	QY	QY	QY
QY	351	QY	QY	QY
QY	352	QY	QY	QY
QY	353	QY	QY	QY
QY	354	QY	QY	QY
QY	355	QY	QY	QY
QY	356	QY	QY	QY
QY	357	QY	QY	QY
QY	358	QY	QY	QY
QY	359	QY	QY	QY
QY	360	QY	QY	QY
QY	361	QY	QY	QY
QY	362	QY	QY	QY
QY	363	QY	QY	QY
QY	364	QY	QY	QY
QY	365	QY	QY	QY
QY	366	QY	QY	QY
QY	367	QY	QY	QY
QY	368	QY	QY	QY
QY	369	QY	QY	QY
QY	370	QY	QY	QY
QY	371	QY	QY	QY
QY	372	QY	QY	QY
QY	373	QY	QY	QY
QY	374	QY	QY	QY
QY	375	QY	QY	QY
QY	376	QY	QY	QY
QY	377	QY	QY	QY
QY	378	QY	QY	QY
QY	379	QY	QY	QY
QY	380	QY	QY	QY
QY	381	QY	QY	QY
QY	382	QY	QY	QY
QY	383	QY	QY	QY
QY	384	QY	QY	QY
QY	385	QY	QY	QY
QY	386	QY	QY	QY
QY	387	QY	QY	QY
QY	388	QY	QY	QY
QY	389	QY	QY	QY
QY	390	QY	QY	QY
QY	391	QY	QY	QY
QY	392	QY	QY	QY
QY	393	QY	QY	QY
QY	394	QY	QY	QY
QY	395	QY	QY	QY
QY	396	QY	QY	QY
QY	397	QY	QY	QY
QY	398	QY	QY	QY
QY	399	QY	QY	QY
QY	400	QY	QY	QY
QY	401	QY	QY	QY
QY	402	QY	QY	QY
QY	403	QY	QY	QY
QY	404	QY	QY	QY
QY	405	QY	QY	QY
QY	406	QY	QY	QY
QY	407	QY	QY	QY
QY	408	QY	QY	QY
QY	409	QY	QY	QY
QY	410	QY	QY	QY
QY	411	QY	QY	QY
QY	412	QY	QY	QY
QY	413	QY	QY	QY
QY	414	QY	QY	QY
QY	415	QY	QY	QY
QY	416	QY	QY	QY
QY	417	QY	QY	QY
QY	418	QY	QY	QY
QY	419	QY	QY	QY
QY	420	QY	QY	QY
QY	421	QY	QY	QY
QY	422	QY	QY	QY
QY	423	QY	QY	QY
QY	424	QY	QY	QY
QY	425	QY	QY	QY
QY	426	QY	QY	QY
QY	427	QY	QY	QY
QY	428	QY	QY	QY
QY	429	QY	QY	QY
QY	430	QY	QY	QY
QY	431	QY	QY	QY
QY	432	QY	QY	QY
QY	433	QY	QY	QY
QY	434	QY	QY	QY
QY	435	QY	QY	QY
QY	436	QY	QY	QY
QY	437	QY	QY	QY
QY	438	QY	QY	QY
QY	439	QY	QY	QY
QY	440	QY	QY	QY
QY	441	QY	QY	QY
QY	442	QY	QY	QY
QY	443	QY	QY	QY
QY	444	QY	QY	QY
QY	445	QY	QY	QY
QY	446	QY	QY	QY
QY	447	QY	QY	QY
QY	448	QY	QY	

Qy	1186	SISRM-----PIEDSSPEKGREAAASGHVIECKSGHI-----	1211
Db	1985	slslsldidannkenepiketeppdsggepskqasgyapksfhvtedtpvcfsrnsul	2044
Qy	1220	--LSYDN-----LKNAREGTRSPRTAHETSLKRSYESVEGNLKGCMRESVPVSAPL	1269
Db	2045	sslsidseddlqecissampkkpsrklkgdnekhspmmgglgedtlldlkidgrpd	2104
Qy	1270	EGLICRALPRGSPHDLKERTVLSGSIOMGTPRATTESFEDGLKYPKOIKRESPIRAFE	1329
Db	2105	se---hqlspdsenfakalgansivsslhqaaaaa-----lsrqassdsllslk	2156
Qy	1330	GAITKGPYDGIITIKEMGRSHEIPRODILTOESRKEPVEVQSTRPIEGSISQGTPIK	1389
Db	2157	sglsisgpf-----hlpdqeekpftsnkprllk-----pwekstletkk	2197
Qy	1390	FDNNS-QQSAIKHNKVASLITGP-----SKLSRGM-PPLEIVPENIKVVERKYEDVRAGET	1443
Db	2198	ieseskgikgkvykslitgkvrnselsgmqkplq---annpmsisrg-----	2244
Qy	1444	VRSHRHSVSSGSPVLRSLTLEHAPRAQLSPGIYDTSARRTPVSYONTMSRGSPMNRTS	1503
Db	2245	-----rtnlh-----ipgy-----rnssestspvskgp	2268
Qy	1504	DVTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDPVVSHPSPOPHRGSPAGEVYW	1561
Db	2269	plktpaekspsegqtattsprgakpsvkseisparqtsqlgssgkapsrgsr-----	2322
Qy	1562	SHUPTOLDPAMPFHRALDPAAAYLQROLSTPPGYPQSOXYLYAMENRQFILNDYITSQ	1621
Db	2323	-----dstprpaqpl-----	2336
Qy	1622	QMOVNLRPDVA---RGLSPREPGLPLPATRGIIDLNPMPPTILVPHPGTSTPPMDRI	1678
Db	2337	pigpggnslspgrngisppnk-----lsqlprt-sspstastks9sgkm	2381
Qy	1679	TYIPGTQITPPRPYNSASMSPCHPTHAAAAASABERERERE-----KERER	1726
Db	2382	sy-----tspggrqmsqqnlt--kgtlgsknassiprsesaskglnqmnngngankkvel	2433
Qy	1727	ERTAAASSDLYLPGSEQPRGSGHYVRSPSPSVRTQETMLQORPSPVQGTNGTSTVITP	1786
Db	2434	srmsstks-----gsed-----rserpvlrvstfkaps-----	2466
Qy	1787	LDPATAQLRIMPLPAGGSPSISQGLPASRVNTAADALAALVAAAAPQMDVSKTESHEA	1846
Db	2467	--ptlr-rkleesastesis---psrpastrsqatpvlpslpdmsls-thssvqag	2519
Qy	1847	A--RLEENLRSAVAASEOQOLEQKTVLEVRKSVQCLYTSSAFPSGPKPQPHUSSVYSEAG	1904
Db	2520	gwrklppln---sptleyndgrpakrhdiar-----shsespsripnrgstwtkrehs	2569
Qy	1905	KDKGPPPKSYEBELTRCKTTITAAAN-----FIDVLIITROIASDKOARERGSSSD	1956
Db	2570	khasslp--rvstwrtrtgsessilssasesseksadekhvnsigtklsgkenvsaqg	2627
Qy	1957	SSSSLSLSHRYE-TPSDAIEIVSPASPAPPOKLTQYQYEVVKANQAFNDPTROYEGLH	2015
Db	2628	twrkikeneifspntstqvtvsgatngaeskltliymapavsk---tedvwvriedcpin	2684
Qy	2016	HYRPOQESPSPOQLLP-----SSOAEHGQVPRPHRLITLADHI	2055
Db	2685	npr---sgrsptgntppvidsvsekanpnikdskdnqakqnvngsvpm--rtvglenrl	2739
Qy	2056	CQIITQDFARNQVSSOTPOQPTF--STFGNPSALVST---PVRTKTSNRYSPESQAQSV	2110
Db	2740	tsfiqvd-----apdqkteikpgqnpvpsvsetnespivert-----pfssssss	2785
Qy	2111	HHQRP-----GSRVSPENLVDKSRGSRPKGSPERSHVSSEPIEYISPPQVPVYH--EKQDS	2164
Db	2786	khsspsqttvaarvtpfnv-----nporkssadatsarsqip-tvpnnntkrrds	2835



## RESULT 17

AAR58634  
ID AAR58634 standard; Protein; 2843 AA.

AC AAR58634;

DT 21-JUN-1995 (first entry)

DE Adenomatous polyposis coli protein (APC).

DE Adenomatous polyposis coli protein gene;

KW familial adenomatous polyposis; colorectal tumor; adenoma.

XX Homo sapiens.

XX W09421814-A.

XX 29-SEP-1994.

XX 21-MAR-1994; 94WO-US02987.

XX 19-MAR-1993; 93US-0034850.

XX (UJYO) UNIV JOHNS HOPKINS.

XX Hill DE, Johnson KA, Kinzler KW, Vogelstein B;

XX WPI; 1994-317033/39.

XX Q-PSDB: Q70633.

XX Antibodies to adenomatous polyposis coli protein - are used for

XX detecting mutations in the APC gene for predicting

XX pre-disposition to cancer, partic. colon cancer

XX Claim 1; Page 46; 81pp; English.

XX Antibodies prepared to this protein are used for detecting mutations

XX in the APC gene for predicting pre-disposition to cancer, particularly

XX colon cancer. The Abs can detect germ line or somatic mutations

XX indicating a predisposition to colon cancer and possibly gastric,

XX oesophageal, pancreatic or small cell lung cancers.

XX Sequence 2843 AA;

XX Query Match 2.98; Score 366; DB 15; Length 2843;

XX Best Local Similarity 19.48; Pred. No. 2.4e-11;

XX Matches 418; Conservative 280; Mismatches 798; Indels 664; Gaps 98;

Qy 347 PEIRKQEQOE-RFQRVGQAGLSATARSEH-----EISEIDGLSEQENNEK 395

Db 981 psiesyeddeskcscvgyqpadlahkhsanhdndgldtldpinyiskysdeqlnsgr 1040

Qy 396 QMRQLSVIPMMFDAQRRVRFKFINNGLMEDPMKYKDQPMNYMTDHE--KEIFKDKFI 453

Db 1041 q-----spsqnerwarpkhileidkqsegrqsrngstpyvtestddkhl 1087

Qy 454 QHPKNFGLIASYLERKSPVDCVLYLYLTKKKENYKALVRRNYKRRGNQOIARPSQEEK 513

Db 1088 kfqphfg-----qqecvspy-----rsrangsetnrvngnhgnqvsgslcqe- 1132

Qy 514 VEEKEDEKAETK--EEEKDEEKEDEKEDSKENTKEKDKIDGTAEETEEREQATPRG 571

Db 1133 -dyeddktpnyserysyeeehheerptnysikyn-eekrhvdpidsykskylatdipss 1190

Qy 572 R-----KTANSQGRKGRITRSMWNEAAAAS-----AAAAATEEPPPPPPP 614

Db 1191 qkqsfstkskssgsgsktehmssssststpsnknqnlhpssaqsrsgqpkkaatck 1250

Qy 615 PEPITSTPEVET-----SRWT-----EEEMEVAKKGLVEHGRNWAATAKMA--- 653

Db 1251 vssinqetigtvcyvedtpicfrcssissaaedgicnqttqeadsantlqlaeikgk 1310

Qy 654 VGTKSE-----AOCKNFYFNKRRHNLNLLQOHH-----OQTSTR 688  
Db 1311 igrsaedpvsevpavsbprtkksrlqgsslsesahkavfespqakspsksgaqtpk 1370  
Qy 689 KPREE-----RDVSQCSVASTVSAQEDIEDIASNEENEPEDSEVAVKPS--DSPEN 740  
Db 1371 spbehvqetplmfsrcstsv-ssidsfesrslaasvqepcsgmlsgilspdlpdsppq 1429  
Qy 741 A--TSRGNTEPAVELEPTETAPSTPSLAVPSTKPADESVETQVNDISIAETAEQMDV 798  
Db 1430 tmpsrsktppp---ppqtaqtrelpnkaptakresgpkgaavnaavq-rvqvlpda 1485  
Qy 799 DQOHSAAE---GSVC-----DPPATKADSVDEVR---VPEH-HASKVEGDN 840  
Db 1486 dtllhfatestpdgfcscsslsalsldpfigk---dvelrmpvqendngneteseq 1541  
Qy 841 TKE--RDLDRASEKVEPRDELVAQOINAORPEQSDNSDSATCSADEVD----- 890  
Db 1542 pkesnengekaektidsekdl-----ddsd-----dddieileecilis 1581  
Qy 891 -----GEPERQRMF-----PMDSKPSLLNPTGSIILVSSPLKPNPLDLQLOHRAVI 937  
Db 1582 amptksrkgkpaqtaslkpppvarkpsqi-pvykllpsq-----nrl---qpqkhvst 1633  
Qy 938 P---PMVSC---TPCNIPICTPVSGYAL-----YORHIKAMHE 969  
Db 1634 pgdmpvrycvegtptinfstatsldtiesppnelaagevrggaqggefekrdtipte 1693  
Qy 970 SAL-----LEEQRQREQIDLECRSSTSPCGTSKSPNREWEVL----- 1007  
Db 1694 grstdeaggktssvtipeldnkaeegdilaecinsampkgkshkpfvkkimdgqgqa 1753  
Qy 1008 ---QPAPHQLITNLPEGVRLTTRTPRPPLI-----P 1038  
Db 1754 sassaapnk---nldgkxkxtpvkvipqnteytrvrknadsknlnaervfsdndk 1810  
Qy 1039 SSKTIVASEKPSF-----IMGGST-----OQTGTYLTSHNOASTQETP 1079  
Db 1811 skkqlknmskdfndklnpnedrvrgsfafdsphhytpiegt-p-ycfsrnd----- 1860  
Qy 1080 KPSVGSISLGLPROQESAKSATLPYIKOEFSRPSQNSQPEGLLVRAQHEGVVGTAGAI 1139  
Db 1861 --slsldfdddvdlsrekaelrkakenkesakvtshte--ltsnqgsa---nktqai 1913  
Qy 1140 QEGSITRGTPTSKISVESIPSLRGSITQGTALPOTGTPTEALVK-----G 1185  
Db 1914 akqplnrgqpkpilkqgs-----tfpqskdipdrgaatdekigqnfalentaipvcfshns 1967  
Qy 1186 SISRM-----PIEDSSPEKGRGEEAASKGHVYEGKSGHI----- 1219  
Db 1968 slsalsldidqennkenepiketepdsqgspkqasgyapakfhvedtpvcfsrns 2027  
Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLKRYSVESVEGNIKQGMRESVPSAPL 1269  
Db 2028 sslsldseddlilqecissampkksrllkgdneksprnmnggilgedtldldkldgrpd 2087  
Qy 1270 EGLICRALPGCSPHSDLKERTVLSGSIHQGTPTATTFESFEDGLKVPKQIKRESPIRAFE 1329  
Db 2088 se---hglspdsenfwdkalqegansivsslhqaaaac-----lsrqassdsdsilsk 2139  
Qy 1330 GAITKGPYDGTITTIKEMGRSIEHPQDILTQESRKTPEVQSTRPIEISISQGTPIK 1389  
Db 2140 sgslgpf-----hltqdeekpftsnkgprllk-----pgekstlekk 2180  
Qy 1390 PDNNS-GQSAIKHNKSLITGP-----SKLSRGM-PPLEIVPENIKVVERGYEDVKAGET 1443  
Db 2181 ieseskgikggkkykslitgkvrnsseisgmkgplq---anmpsirsrg----- 2227  
Qy 1444 VRSRHTSVSSGSPSVLSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMWSRGSPMMNRTS 1503  
Db 2228 -----rtmih-----ipgv-----rnssstspvskkgp 2251











OS Homo sapiens.  
 XX US5352775-A.  
 XX PD 04-OCT-1994.  
 XX PF 08-AUG-1991; 91US-0741940.  
 XX PR 16-JAN-1991; 91GB-0000962.  
 XX PR 16-JAN-1991; 91GB-0000963.  
 XX PR 16-JAN-1991; 91GB-0000974.  
 XX PR 16-JAN-1991; 91GB-0000975.  
 XX (CANC-) CANCER INST.  
 PA (ICIL) IMPERIAL CHEM IND.  
 PA (UJXO) UNIV. JOHNS HOPKINS.  
 PA (UTAH) UNIV UTAH.  
 XX Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;  
 PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;  
 PI Vogelstein B, White RL;  
 XX WPI; 1994-316233/39.  
 DR N-PSDB; AAQ72297.  
 XX New human adenomatous polyposis coli DNA encoding tumour  
 PT repressor - and derived primers and probes for diagnosis,  
 PT prognosis and treatment of cancer  
 XX Claim 3; Columns 71-84; 113pp; English.  
 XX AAQ72297 is a cDNA isolated from the human adenomatous polyposis  
 CC coli (APC) gene, it encodes the tumour repressors described in AAR63507  
 CC and AAR63508. Determination of alterations in APC or its expression  
 CC products, can be used for the diagnosis and prognosis of cancer e.g.  
 CC colorectal, lung and breast tumours; and for determining predisposition  
 CC to certain cancers such as familial adenomatous polyposis (FAP) and  
 CC Gardner's syndrome. The wild type APC gene (or a part of it) can be  
 CC used therapeutically to restore gene function, while primers and probes  
 CC derived from the cDNA (AAQ72333-400 and AAQ72541-568) can be used to  
 CC detect mutations. Also APC proteins or analogues can be administered to  
 CC compensate for a defective gene, and epithelial cells, or transgenic  
 CC animals carrying a mutated APC allele are useful for detecting  
 CC therapeutic agents able to suppress tumorigenesis.  
 XX Sequence 2842 AA;  
 SQ

Query Match 2.9%; Score 365; DB 15; Length 2842;  
 Best Local Similarity 19.0%; Pred. No. 2,7e-11;  
 Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKQROQE-RQQRVGQRCAGISANTARSEH-----BISEIIDLSEQENNEK 395  
 DB 980 psiesyseddeskcsyqypadiahkhsanhdndgdeldtpnyslkysdeqlnsgr 1039  
 QY 396 QMRQLSVIPPMFADQORRVFINNGLMEDPMKVYKDQPMNVWTDHE--KEIFKDKFI 453  
 DB 1040 q-----spqnerwarphkhiiedeikqseqqrnsqntypvytestdkhl 1086  
 QY 454 QHPKNFGLIASYLERKSVDPCLVYYTKKNENYKALVRRNYGRRGRNQOIARPSOEK 513  
 DB 1087 kfqbhf-----q-qecvspy-----rsrgangsetnrvsnhngnqvsgslqce- 1131  
 QY 514 VEEKEEDKAETEK--EEKKDEEKEDSKENTKEKDKIDGTAEETEERQANPRG 571  
 DB 1132 -ddyddedkptnysryseeqneeeerptnysikyn-eehrhvdqpdidslykatdipss 1189  
 QY 572 R-----KTANSQRRRGRTIRSNTEAAAS-----AAAAATEEPPLPLPP 614  
 DB 1190 qkqsfstksqsgsksthemsssentstpsnaktqnqlhpsaqrsgqpkaatck 1249  
 QY 615 PEPISSTEPVET-----SRWTEEMEVAKKGLVEHGRN-----WAAIAKM 653

DB 1250 vssinqetiqtvcvedtpicfsrcsslsslaaedeigcncqtqeadsantlqiaek 1309  
 QY 654 VGTKSE-----AQCKNFYFNYKRRHLDNLQOHK-----QKTSR 688  
 DB 1310 igtrsaedpvsevpavsqhprtkssrlqgsslsesarhkavsfssgakspsksgagtpk 1369  
 QY 689 KPREE-----RDVSOCEVASTVSAOEDIEDASNEENPEDESEVAVKPE--DSPEN 740  
 DB 1370 sppehyvqetplmfrctsv--ssidsfesrsiasvsvepcsgmvsgliispsdlpdpqg 1428  
 QY 741 A--TSRGNTEPAVELETTTAPSTSPSLAVPSTKPAEDESVEVQVNDISIAETAQMDV 798  
 DB 1429 tmppsrsktppp---ppqtaqtkevvpknkaptakresgpkqaavnaavq-rvqvlpda 1484  
 QY 799 DQEHSAEE-----DPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840  
 DB 1485 dtllhfatespdgfsccsslsalsldepfigk----dveirimpvqgndngneteseq 1540  
 QY 841 TKE--RDLDRASEKVEPRDEDLVVAQINAQRPQPQSDNDSSATCSADEVD----- 890  
 DB 1541 pkesnenqeakektidsekdl-----ddsd-----ddtleileecilis 1580  
 QY 891 ---GEPERQRMFPMDSKPSLLNPTGSLVSSPLKPNPLD---LP-----OLQHRAAVIP 938  
 DB 1581 amptkssrkakpaqtasklpppvar-----kpsqlpvylkllpsqnrllpqkhvstfp 1633  
 QY 939 ---PMVSC---TPCNIPITGTPVSGYAL-----YORHIAKHUES 970  
 DB 1634 gdmprvycvegtplnfstatsldltiesppnelaagevrgvgaqsgefekrdtiptg 1693  
 QY 971 AL-----LEBORQEQIDLECRSSSPCGTSKSPNREWEVL----- 1007  
 DB 1694 rstdeaggktsvvtipelddnkaeedilaecinsampgkshkpfvrvkkmidqvqas 1753  
 QY 1008 --QPAPHQLITNPEGVRLPTTRTPPPPLI-----PS 1039  
 DB 1754 assapnk---nqidgkkktpspvpipqpqrtrvrknadsknnlnaervsfndkds 1810  
 QY 1040 SKTIVASEKSF-----IMGGSIS-----QGTPTGYLTSHNQASTQETPK 1080  
 DB 1811 kqnlkanskdfndklpnnedrvrgsfafdsphhytlegtp--ycfsrd----- 1859  
 QY 1081 PSVGSISLGLPRQOESAKSATPIYKQEEFSPRSQNSQPEGLLVRAQHEGVVRCAGIAI 1140  
 DB 1860 -sissldfddddvdsrekaelrkakeneseakvtshte--ltsnqgsa---nktqaa 1913  
 QY 1141 EGSTRGTPTSKISVESIPSLRGSIQTGTPALPOTGTPTEALVK-----GS 1186  
 DB 1914 kqpinrggpkpilkqs-----tfpqsskdipdrgaatdeklqfalentpvcfshms 1967  
 QY 1187 ISRM-----PIEDSSPEKGREEAASKGHVIECKSGHI----- 1219  
 DB 1968 lssisdidqennnkenepiketepdsqgpekpqasgypaksfhvedtpvcfarnssls 2027  
 QY 1220 -LSYDN-----IKNARECTRSPTAHEISLRSVESVEGNIKQGMRESVPVSAPLE 1270  
 DB 2028 slsidseddllqecissampkkkpsrllkgdnekhspnmngilgedltldkdiqrps 2087  
 QY 1271 GLICRALPRGSPHDLKERTVLSGSIHQGPTRATESFEDGLKYPQIKRESPIRAPEG 1330  
 DB 2088 e---hglspsdsenfdkaiqegansivsslhqaaaac-----lsrqassdsdlsilks 2139  
 QY 1331 AITKGKPYDGTITIKEMGRSIEHPRDILTQESRKPTVEVQSTRPIEGSISQGTPIKF 1390  
 DB 2140 gislgspf-----hltpdqeekftsnkgrlik-----pgekstlektki 2180  
 QY 1391 DNNS-GOSAIKHNKYSKITGP-----SKLSRGM-PPLIEIVPENIKVPRGYEDYKAGETV 1444  
 DB 2181 eseskgikgkvykslitgkvrsnseisgmkplq---anmpsirsrg----- 2226  
 QY 1445 RSRTSVVSSGSPVSRSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMTMSRPMNRSTD 1504



```
Db 2227 -----rtmih-----ipgv-----rnssstspvskkqpp 2251
Qy 1505 VTIPNKNSTHERKSTLTPT-QRESIPAK-SPVPGVDVWVSHSPDPHHRGSTAGEVYWS 1562
Db 2252 lktbaskspsegqtattisprgakpsvkselprqtsiqgsskapsrsgs----- 2304
Qy 1563 HLPQLDPAMPFHRALDPAAAYLFQROLSTPTGPGSYQYLYAMENTROTILNDYITSQQ 1622
Db 2305 -----dtparpadqpl-----srp 2319
Qy 1623 MQVNLRPDVA---RGLSPREOPLGLPYPATRGIIIDLTMNPPTILVPHPGG'STPMDRIT 1679
Db 2320 iqsggrnisprgriqppnk-----lsqilpt-spsstaskssgsgkms 2364
Qy 1660 YIPGTQITFPPRPNYSMSMGPHPTHAAAAAERERERE-----KERERE 1727
Db 2365 y-----tspgrqmsqult--kqtglsknassipresaskglnqnmngngankkvls 2416
Qy 1728 RIAAASDLILRPGSEQPRGSHGVVRSPSPSVTQETMLQORPSVFGTNGTSVITPL 1787
Db 2417 rmsstkss-----gseqd-----rserpvivrgstfikeaps----- 2448
Qy 1788 DPTAQLRIMPLPAGGPGSTISQGLPASRYNTAADALAALVDAASAPQMDVSKTSKHEAA 1847
Db 2449 -ptlr-kieesafesis---psirpasptrsqatpvlspisdmsls-thsvqagg 2502
Qy 1848 --RLEENLRSAVSEQQOLEQKTLVEKRSVQCLYTSSAFFSGKPPQHSVVVSEAGK 1905
Db 2503 wrklppnl---sptieyndgrpakrhdia-----shsesparlpinrgtwkrehs 2552
Qy 1906 DKGPPPKRYEEELTRGKTITTAAN-----FIDVIITROIASDKDARERSQSSDS 1957
Db 2553 hssalp--rvatwrtgsssilassessekaksedekhvnsisgtkqskengvsaagt 2610
Qy 1958 SSSLSHRYE-TPSDATEIVISPASSPAPPQEKLOTYOPEVVKANQAENDP'ROVEGLHH 2016
Db 2611 wrkikeneftstntstqtvssgaangaeskliymgavask-----tedvvrriedcpinn 2667
Qy 2017 YRPOQESFPOOQLPP-----SSQAEFGMGQVPRTHRLITLADHIC 2056
Db 2668 pr---sgrsptgnppvidsvsekanpnikskdnqakqngvngsvpm--rtvglenrln 2722
Qy 2057 QIITQDFARNQVSSQTP-QQPTSTFQNSPSALYS-TPVRKTKSNRYSPEAQOAVHHQR 2114
Db 2723 sflqvdapdqgtktpqgnnpvpvsetnessivertpfssssskshspsgtva----- 2777
Qy 2115 PGRSVSPENLVKSGRSPGKSPERSHVSSEPEYPIPPQVPVVH--EKQDS 2164
Db 2778 --arvtfny-----npsprkssadtsarpsqip-tpvnnntkkrds 2817

RESULT 20
AAW76140
ID AAW76140 standard; Protein; 2843 AA.
XX
AC AAW76140;
XX
XX 23-NOV-1998 (first entry)
DT
DE Human APC protein #1.
XX
XX Familial adenomatous polyposis coli; APC; tumour suppressor; therapy;
KW chromosome 5q21; tumorigenesis; retinoblastoma; colorectal tumour;
KW FAP; Gardner's Syndrome; GS; predisposition.
XX
OS Homo sapiens.
XX
XX US578366-A.
XX
XX 21-JUL-1998.
PD
XX
XX 25-MAY-1995; 95US-0452655.
XX
```

```
PR 16-JAN-1991; 91GB-0000975.
PR 16-JAN-1991; 91GB-0000962.
PR 16-JAN-1991; 91GB-0000963.
PR 16-JAN-1991; 91GB-0000974.
XX
PA (CANC-) CANCER INST.
PA (UYJO ) UNIV JOHNS HOPKINS.
PA (UTAH ) UNIV UTAH.
PA (ZENE ) ZENECA PHARM.
XX
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ;
PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A;
PI Vogelstein B, White RL;
XX
DR WPI; 1998-427100/36.
DR N-PSDB; AAV56447.
XX
PT Adenomatous polyposis coli protein - useful in the treatment of
PT cancers associated with mutation(s) on human chromosome 5q21
XX
PS Disclosure; Column 41-54; 102pp; English.
XX
CC This sequence represents a human familial adenomatous polyposis coli
CC (APC) protein from clone DP2.5. The gene for the protein is present on
CC human chromosome 5q21 and is also referred to as adenomatous polyposis
CC coli gene. It is a tumour suppressor gene, and mutations in this gene
CC have been associated with tumorigenesis in retinoblastoma and colorectal
CC tumours, and especially familial adenomatous polyposis (FAP) and
CC Gardner's Syndrome (GS). The protein can be used in therapy to replace
CC lack of native functional protein and the nucleic acids can be used for
CC gene therapy. The nucleic acids that encode them can also be used as
CC probes and primers in detection of the cancers and predisposition to it.
XX
SQ Sequence 2843 AA;

Query Match 2.9%; Score 365; DB 19; Length 2843;
Best Local Similarity 19.0%; Pred. No. 2.7e-11;
Matches 409; Conservative 286; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQREQE-RQRVQRCAGLSATARSEH-----EISEIIDGLSEQENNEK 395
Db 981 pslesyeddeskcscsygypadiahkhsanhdmdndgeidtpinyiskysdedqnsgr 1040
Qy 396 QMRQLSVIPPMFMDAEQRRVFINNGLMEDPMKYKDRQFMNVTDHE--KEIFKDFI 453
Db 1041 q-----spsqnerwarpkhiedeikseqqrngsttypvtestddkhl 1087
Qy 454 QHPKNFGLIASYLERKSVDPCLVLYYYLTKMKNYKALVRRNYKRRNQIARPSQEEK 513
Db 1088 kfphfg-----qgecvsp-----rsrgangsetnrvsnhgnqnvsgslcqe- 1132
Qy 514 VEEKEEDKAETKK--EEKKDEEKDEKEDSKENTKEKIDGTAETETEREQATRG 571
Db 1133 -ddyeddkptnyseryseehheerptnysikyn-eehrhvdqpidysikyatdipss 1190
Qy 572 R-----KTANSQRRKGRITRSMTEAAAS-----AAAAAATPEPPPLPPP 614
Db 1191 qkqsfsksssgskstehmsssentstpsnakrnqlhpsaqsrsgqkaactk 1250
Qy 615 PEPISTEPVET-----SRWTEEMEVAKKGLVEHGRN-----WAAIARM 653
Db 1251 vssinqetiqlcyvedpicfscsslsllsaedeigcncqtqeadsantliaelke 1310
Qy 654 VGTKE-----AQCKNFYFNYKRRHLDNLLQOHK-----QKTSR 688
Db 1311 igtrsaedpvsevpavsqhprtksrliqgsslsasathkavefssgakspsksgaqtpk 1370
Qy 689 KPREE-----RDYSQCESVASTVSAQDEDEIDIASNEENPEDSEVEAVKPSE--DSPEN 740
Db 1371 sppenyvqetplmfsrctsv--ssldsfesrslasvqpcsgmvsnglispdlpdpqpg 1429
Qy 741 A--TSRGNTPEAVELEPTTETAPSTSPSLAVPSTKPAEDSVETQVNDISAETAQMDV 798
```







PI Vogelstein B, White RL;

XX WPI; 1998-427100/36.

XX Adenomatous polyposis coli protein - useful in the treatment of  
XX cancers associated with mutation(s) on human chromosome 5q21

PS Disclosure; Column 63-78; 102pp; English.

XX This sequence represents a human familial adenomatous polyposis coli  
XX (APC) protein isolated from 87 cDNA clones. The gene for the protein is  
XX present on human chromosome 5q21 and is also referred to as adenomatous  
XX polyposis coli gene. It is a tumour suppressor gene, and mutations in  
XX this gene have been associated with tumorigenesis in retinoblastoma and  
XX colorectal tumours, and especially with familial adenomatous polyposis (FAP)  
XX and Gardner's Syndrome (GS). The protein can be used in therapy to  
XX replace lack of native functional protein and the nucleic acids can be  
XX used for gene therapy. The nucleic acids that encode them can also be  
XX used as probes and primers in detection of the cancers and predisposition  
XX to it.

XX Sequence 2843 AA;

Query Match 2.98; Score 365; DB 19; Length 2843;

Best Local Similarity 19.08; Pred. No. 2.7e-11;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKQEQE-RFQRVQRCAGLSATIASEH-----EISELIDGLSQENNEK 395  
DB 981 pslesyeddeskfcsygyqpadlahlhanhmdndgeltdtpnlskysdeqlnsgr 1040  
QY 396 QMRQSLVPPMFDQORRVKFINNMGLMDPMKVKYDRQFMNVTMDHE--KEIFKDKFI 453  
DB 1041 q-----spqnerwarphkiedekqseqgrndqsttppyytestdkhl 1087  
QY 454 QHPKNFGLIASYLERKVPDCLVLYYTKKNENYKALVRRNYKRRGRNOQIARPSQEK 513  
DB 1088 kfqbhfg-----qqecvspy-----tsrgangsetnrvnshngingnvsqslcqe- 1132  
QY 514 VEEKEDEKAEKTEKK--EEKKDEEKEDKEDSKENKTEKDKIDGTAEETEEQATPRG 571  
DB 1133 -ddyeddkptnysyeeeqheerptnyslkyn-eehrhwdpdlidslykatdlps 1190  
QY 572 R-----KTANSQRRKGRITRSMTEAAAS-----AAAAAYEPPPPPLPPP 614  
DB 1191 qkqsfksksgsgsktelhmsassentstpsnaktqnglhpsaqsrggqkaatek 1250  
QY 615 PEPISPEVET-----SRWTEEMEVAKKGLVEHGRN-----WAAATAM 653  
DB 1251 vssinqetldtqcvetpdcfrsslsissaeedlgcnqtteadsantlqlaeikek 1310  
QY 654 VGNKSE-----AQCNFNFNKRRNLDNLLQHK-----QKTSR 688  
DB 1311 igtrsaedpvpavsqhprtkssrlqgsslsesesarhkavfssgskapskgactpk 1370  
QY 689 KPREE-----RDVSOCEVASTVSAQEDDEIASNEENPEDESEAVKPSE--DSPEN 740  
DB 1371 sppehyqetplmfrctsv-saldsfesrsiassvqsepcsmgmvgsiispsdlpdpqg 1429  
QY 741 A--TSRGNTPEAVELEPTTAPSTPSLAVPTKPAEDESVEQVNDSDISAETAQOMDV 798  
DB 1430 tmpsrsktppp--ppqtaqtrevpknkaptkaekresgpkqaavnaavq-rvqvlpda 1485  
QY 799 DQGEHSAAE-----GSVC-----DPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840  
DB 1486 dtllhfatepdgfcscsslsalsidepfik-----dveirmpvqndngneteseq 1541  
QY 841 TKE--RDLDRASEKVEPRDLDVVAQINQAQPEPQSDNDSSATCSADEVD-----890  
DB 1542 pkesnenqekaektldsekdl-----ddsd-----dddielleecils 1581  
QY 891 ---GEPEQRMFPMDSKPSLLNPTGSLVSSPLKPNPLD-----LP-----QLQHRAAVIP 938

DB 1582 amptkserrakkpaqtasklpppvar-----kpsqlpvkylipsqnlqpkhvsftp 1634  
QY 939 ---PMWSC---TPCNIPIGTPVSGYAL-----YQRHKAMHES 970  
DB 1635 gdmprvyvegtpinfstatisldltiespnlaeagevrgaagsefektldipteg 1694  
QY 971 AL-----LEEQRQREQIDLECRSSSPCTGTSKSPNREWEVL-----1007  
DB 1695 rstdeagggktsstvipelddnkaeegdilaecinsampkshkpfvkkimqvgqas 1754  
QY 1008 --QAPAHQLITNLPEGVNLPTTRTRPPPLI-----PS 1039  
DB 1755 assaapnk---nqidgkktktpvskpqpnteyrtvrknadsknnlnaervsfndkds 1811  
QY 1040 SKTVVASEKPSF-----IMGSGTIS-----OGTPGYLTSNHNQASYQTETPK 1080  
DB 1812 kknlnmskdfndklpnnedrvrgsfafdsphhytiegtp--ycfsrnd-----1860  
QY 1081 PSVGSISILGLPRQOESAKSATLPYIKQEEFSPRSQNSQPEGLLVRAQHEGVVGTAGAIQ 1140  
DB 1861 -slsldfddddvdsrekaelrkakeneseakvtshte--ltsnqgsa---nktqai 1914  
QY 1141 EGSTRGTPTSKISVESIPSIRGSIQTGTPALPOTGTPTEALVK-----GS 1186  
DB 1915 kqpinrgqpkilqks-----tfpqsskdlpdrgaatdeklqnfaientpvcfshns 1968  
QY 1187 ISRM-----PIEDSSPEKGREEAASKGHVIECKSGHI-----1219  
DB 1969 lssldidqennnkenepliketepdsqgskpqasgypaksfhvedtpvcfarnsals 2028  
QY 1220 -LSYDN-----IKNAREGTRPRTAHEISLRSVESVEGNIKQGMRESVPVSAPLE 1270  
DB 2029 slsidseddillqecissampkkkpsrlkgdnekhspnmggilgedltlklkdiqrds 2088  
QY 1271 GLICRALPRGSPHSDLKERTVLSGSIHQGTPTPTATESFEDGLKYPKQIKRESPPIRAFEG 1330  
DB 2089 e---hgispdsenfwaigegansivsslhqaaaaac-----lsrqassdsalsiks 2140  
QY 1331 AITKGKPYDGTITIKEMGRSITHEIPRODILTQESRKTPVEVQSTRPIIEGSIQGTPIKF 1390  
DB 2141 gislgspfi-----hltpdqeekpftsnkgrprlik-----pgekstletkkl 2181  
QY 1391 DNNS-GOSAIKHNKVLITGP-----SKLSRGM-PPLETPENIKWVERGKYEDVKAGTV 1444  
DB 2182 esekgikgkkykslitkvrnseniseigmkplq---anmpsiarg-----2227  
QY 1445 RSRTSVSVSGPSVLRSTLHEAPKAQLSPGIYDDTSAARTPVSYQNTMSRGSPPMNRSTD 1504  
DB 2228 -----rtmih-----ipgv-----rnsststspvskgpp 2252  
QY 1505 VTTPPNKSTNHERKSTLTPT-QRESIPAK-SPVPQVDPVVSHPDFPHHRCSTAGEVYWS 1562  
DB 2253 lktpasksegqtattsprgakpsvkselspvarqtsqiggskapsragsr-----2305  
QY 1563 HLTQOLDPAMPFHRALDPAALFQRLSPTPCYPOYOLYAMENTROTILNDYITSOQ 1622  
DB 2306 -----dstprpaqqpl-----srp 2320  
QY 1623 MQVNLRPDVA---RGLSPREQPLGLPYATRGIIIDLTMPPPTILVPHPGGTSTPMDRIT 1679  
DB 2321 iqpggrnispgrngisppnk-----lsqprt-ssptaastkssgskms 2365  
QY 1680 YIPGTQITFFPRPNNSMSFGHPHTLAAAASAEERERERE-----KEERE 1727  
DB 2366 y-----tspggrmsqqlt--kgtglsknassiprsesaskglnmngnngankkvels 2417  
QY 1728 RIAAASDLYLRPGSEQPGRPGSHGVYVRSPSVRTQETMLQORPVSFQGTNGTNGTITPL 1787  
DB 2418 rmstks-----gsead-----rserpvlrvgtstfikeaps-----2449  
QY 1788 DPTAQLRIMPLPAGGFSISQGLPASRYNTAADALALVDAASAPQMDVSKTESKEHAA 1847



Db 2450 -ptlr-rkleeasfesls---psrpsatrsqatpvlpslpdmsls-thssvqag 2503  
QY 1848 --RLEENLRSAVSEQQLEQKLEVEKRSVQCLYTSSAFPSGKQPPOPHSSVYSEAGK 1905  
Db 2504 wrklppnl---sptleyndgrpakrhdiar-----shsesprlpinrgtwtkrehs 2553  
QY 1906 DKGPPPKRYEELRGTGTTAAN-----FIDVILTRQIASDKDAREGSQSDS 1957  
Db 2554 hssslp--rvstwtgsssilssassekaksekdekhvnsigtqskengvsaagt 2611  
QY 1958 SSSLSHRYE--TPSDAIEVISPASPAPQKLOQYEVVKANQAEENDPRQYEGPLHH 2016  
Db 2612 wrkikenesptnsgtvsygatngaesktliygmapavsk-----tedvwrwiedcpinn 2668  
QY 2017 YRQOESPSQQQLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056  
Db 2669 pr---sgrsptgntppvidsvsekanpnikskdnqakngvngsvpm--rtvglenrlin 2723  
QY 2057 QIITQDFARNQVSSQTP-QQPPTSTFQNSALVS-TPVRTKTGNRYSPESQAOVHHQR 2114  
Db 2724 sficvdpdqgkteikpgnnpvpvsetnessivertpfssssskhspsgtva----- 2778  
QY 2115 PGRSVSPENLVKSRGSPGSPERSHVSSEPEPISPPQVPPVH--EKQDS 2164  
Db 2779 --arvtpfny-----npsprkssadtsarpsqip-tpvnmntkrds 2818  
RESULT 22  
AAB23011  
ID AAB23011 standard; Proteins; 2843 AA.  
XX AC AAB23011;  
XX AC  
XX DT 16-JAN-2001 (first entry)  
XX DE Human APC protein (splice variant 1).  
XX KW APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;  
XX KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;  
XX KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;  
XX KW bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;  
XX KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;  
XX KW genetic predisposition; drug screening; DP2.5; splice variant.  
XX OS Homo sapiens.  
XX PN US6114124-A.  
XX PD 05-SEP-2000.  
XX PF 25-MAY-1995; 95US-0450582.  
XX PR 16-JAN-1991; 91GB-0000962.  
XX PR 16-JAN-1991; 91GB-0000963.  
XX PR 16-JAN-1991; 91GB-0000974.  
XX PR 16-JAN-1991; 91GB-0000975.  
XX PR 08-AUG-1991; 91US-0741940.  
XX PR 12-AUG-1994; 94US-0289548.  
XX (ICIL ) IMPERIAL CHEM IND PLC.  
XX (UYJO ) UNIV JOHNS HOPKINS.  
XX (UTAH ) UNIV UTAH.  
XX (CANC-) CANCER INST.  
XX Carlson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R;  
XX Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B;  
XX Hedge PJ;  
XX WPI: 2000-565003/52.  
XX N-PSDB: AAA93449.  
XX Detecting Adenomatous Polyposis Coli (APC) protein in a sample for  
XX diagnosing cancers, involves contacting the sample with antibodies that

PT specifically bind to APC protein and detecting the complex formed -  
XX Claim 1; Fig 3A-C; 125pp; English.  
XX The invention relates to a novel method for detecting Adenomatous  
XX Polyposis Coli (APC) protein in a sample. The method involves  
XX contacting the sample with antibodies which specifically binds to the  
XX 2843 amino acid form of the human APC protein, or to a mutant APC  
XX protein, and detecting an APC-antibody complex. Mutations in the APC  
XX gene play a role in tumorigenesis, indicating that it is a tumour  
XX suppressor gene. It is located on chromosome 5q21, which corresponds to  
XX the FAP (familial adenomatous polyposis) locus. FAP is an autosomal  
XX dominant inherited disease in which affected individuals develop  
XX hundreds to thousands of adenomatous polyps in the colon and rectum,  
XX some of which progress to malignancy. The FAP locus is often found to  
XX be deleted in sporadic (i.e., non-familial) adenomas and carcinomas, and  
XX chromosome 5q deletions have also been observed in tumours of the lung,  
XX breast, colon, rectum, bladder, liver, sarcomas, stomach, and prostate,  
XX and in leukaemias and lymphomas. Although the FAP locus contains  
XX several other genes such as FER, TBL1, TBL2, and MCC, it is thought that  
XX mutations in the APC gene play a key role in the development of FAP and  
XX sporadic tumours. The method is useful for detecting APC protein and its  
XX mutant forms in foetal tissue, placental tissue, amniotic fluid, blood,  
XX serum or a tumour sample. The method is useful for diagnosing or  
XX prognosing neoplastic tissue, for detecting a genetic predisposition to  
XX cancer, for detecting germline and somatic alteration of wild-type APC  
XX genes, and for testing therapeutic agents for the ability to suppress  
XX tumours. The present sequence represents a 2843 amino acid splice  
XX variant of the human APC protein. This variant is more abundant than the  
XX 2742 amino acid variant (AAB23012).

XX Sequence 2843 AA;

Query Match 2.98; Score 365; DB 21; Length 2843;

Best Local Similarity 19.08; Pred. No. 2.7e-11;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PEIRKREQQB-RPQRVGQAGLSATIRSEH-----EISEIDGLSQENNEK 395  
Db 981 psiesyeddeskscsygypadlahkhsanhmddngeldtpinyslkysdeqlnsgr 1040  
QY 396 QMRQLSVIPPMFAEORRVKFINNGLMEDPMKYKQRFPMVWTDHE--KEIFKKKFI 453  
Db 1041 q-----spqnerwarpkhiiledikqseqrgsqrdsttvpvtestdkhl 1087  
QY 454 QHPKNFGLIASYLERKSVPCVLYLYLTKKENYKALVRRNYKRRGRNQOIARPSQEEK 513  
Db 1088 kfqphfg-----qecvspy-----rsrgangsetnrvgshnglnvsgslqce- 1132  
QY 514 VEEREEDKAETK--EKKDEEEDKEDSKENTKEDKIDGTAEETEREQATPRG 571  
Db 1133 -dyeddktptnyseryseseehereptnysikyn-eehrhvdqpidyslkaytdipss 1190  
QY 572 R-----KTANSQGRKGRITRSMTEAAAS-----AAAAATEPEPPPLPPP 614  
Db 1191 qkqsfstskssggsgsktehmssssststpsnkrqnlhpsaagrsqgpkkaatck 1250  
QY 615 PEPISSTEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAATKM 653  
Db 1251 vssinquetigtvcvedtptcfscrossissaeedlgcnqtdqeadsantlqlaelkek 1310  
QY 654 VGTKSE-----AQCKNFYFNKRRHNLNLOQHK-----QKTSR 688  
Db 1311 igtrsaeidpvsavsqhqrptksrriqgsslsesarhkavefssgakspsksagatpk 1370  
QY 689 KPREE-----RDVSSQCESVASTVSAQDEDEIDEASNEENPEDESEVAVKPE--DSPEN 740  
Db 1371 sppehyvqetplmfsrctsv-ssidsfesrsiassvqpcsgmvgiispsdlpdpqpg 1429  
QY 741 A--TSRGNTEPAVELEPTTETAPTSTPSLAVPSTKPAEDESVEVQVNDISAEAEQMDV 798  
Db 1430 tmppsrsktppp---pbqtaqtrevvpknaptakresgpkqaavnaavq-rvqvlpda 1485



Qy	799	DQOEHSAAEE	-----GSVC-----	-----DPPPATKADSVDEVR-----	-----VPEN-HASKVGEDN	840
Db	1486	dtllhrtatestpdgfcscsslsalsldeplqk	-----dveirimpvqendngneteseq	-----	1541	
Qy	841	TKE	-----RDLDRASKEPRDEDLVVAQOINAQPEPOSDSSATCSADEVD	-----	890	
Db	1342	pkeenenqeakaektldsekdl	-----dads-----	-----dddtlelleecils	1581	
Qy	891	-----GEPRQRMFMDSPKLLNPTSGILVSSPLKPNPLD	-----LP-----	-----OLOHRAAVIP	938	
Db	1582	amptkssrakakpaqtaskalpppvar	-----kpsqlpvyklpsqnrlqpqkhvstfp	-----	1634	
Qy	939	-----PWVSC-----TPCNIPITGPVSGVAL	-----	-----YORH1KAMHES	970	
Db	1635	gddmrvyvegtpinfstatsldtiesppnelaagevrggaqsfekrdtipteg	-----	-----	1694	
Qy	971	AL	-----LEEQRQROEQIDLECRSSTSPCGTSKSPNREWL	-----	1007	
Db	1695	rstdeagggktsvtlpelddnkaeegdllaecinsampgkshkpfvrkkmddqvqas	-----	-----	1754	
Qy	1008	--OPAPHLQTLNLPPEGVRLPFRPRPPPLF	-----	-----PS	1039	
Db	1755	assapnk	-----nqlodgkkktpsvkbpqpnteytrvrknadsknlnaervsfndkds	-----	1811	
Qy	1040	SKTTVAASEKPSF	-----IMGSIS-----	-----OGTPGYLTJSHNOASYTQETPK	1080	
Db	1812	kkqnlknnskdndkpinnedrvrgsfatdshphtyplegt	-----ycfsrnd-----	-----	1860	
Qy	1081	PSVGSISLGLPROESAKSATPLYIKQEFSPRSONSQOPELLVRAQHEGVGRGTAGAIQ	-----	-----	1140	
Db	1861	-slssldfddvdlrsrekaerlakenekesekvtshte	-----ltsnqgsa-----	-----nktqgaia	1914	
Qy	1141	EGSTIRCTPTSKISVESIPSLRGSTIQGTPALPQGIPTREALVK	-----	-----GS	1186	
Db	1915	kqpinrgqpkplqqs	-----tfpqaskdlpdrgaatdeklqnfaientpvcfshms	-----	1968	
Qy	1187	ISRM	-----PIEDSSPKGREAAASKGHVIEGKSGHI	-----	1219	
Db	1969	lssldidqennknepiketepdsgdgsqkpsqayapksthvedtpvcfarnssls	-----	-----	2028	
Qy	1220	-LSVDN	-----IKNAREGTRSPRTAHEISLARSYESVEGNIKQGMRESVPVSAPLE	-----	1270	
Db	2029	slsldseddlqlecissampkkksrllkgdneksnrmggilgeditldkldqrpds	-----	-----	2088	
Qy	1271	GLICRALPRGSHSLKERTVLSGSTMQCTPRATTESPEDGLKYPKOIKRESPPIRAFEG	-----	-----	1330	
Db	2089	e---hglspsdsenfukaIQegansivssllhqaataac	-----lsraqssdsdsilks	-----	2140	
Qy	1331	AIYKGPYDGITTIKEMGRSHIHEIPRODILTQESRKTPPEVQSTRPIEGSISQGTPIKF	-----	-----	1390	
Db	2141	gislgspgf	-----hltpdqeeqfptsnkqgprllk	-----pgkestletckkl	2181	
Qy	1391	DNNS-GOSALKHNKSLITGP	-----SKLSRGM-PLEIIVPENIKVPERGYEDVKAGETV	-----	1444	
Db	2182	eseskgikggkkykslitgvrinseslgmqkqplq	-----anmpsigr	-----	2227	
Qy	1445	RSRHTSVSSGSPVSLRSTLHEAPKQLSPGIYDDTSARRTPVSVQNTMTSGSPMMNRTSD	-----	-----	1504	
Db	2228	-----rtmth-----	-----lpgv-----	-----rnsststspvskkgpp	2252	
Qy	1505	VTIPPNKSTNHERKSTLTPT-QRESIPAK-SVPVGVDVWVSHSPDFPHHGRSTAGEYIWS	-----	-----	1562	
Db	2253	lktpaakspsegqtattspgkpsvksealspvarqtsiqgssakapsrsgsr	-----	-----	2305	
Qy	1563	HLPTQLDPAMPFRHALDPAAAYLFORLSPTPGVPSPQYLIYAMENRTQTLINDYITTSOQ	-----	-----	1622	
Db	2306	-----	-----dstpsraqppl-----	-----srp	2320	
Qy	1623	MOYNLRPDVA	-----RGLSPREQPLGLPYPATRGIIDLTMNPPTILVPHPGGTSPPMDRIT	-----	1679	
Db	2321	iqspgrnslspgrnglspbnk	-----lsqilprt-sspatackssgsgkms	-----	2365	

Qy	1680	YIPCTQTITFPFRPNYNASMSPGCHPHLAAAAAERERERE-----KERERE	172
Dd	2366	y-----tspgrqmsqgnit--kgtglsknasiprseasakglnmngnngankkvcls	2417
Qy	1728	RIAAASDLYLRGCSOPGRPGSHGVVRSPSSVTRTQETMLQORPSVFQGTNGTSVTIPL	1787
Dd	2418	lmstks-----gsesd-----rsrpvlvrqsflkeaps-----	2449
Qy	1788	DPTAQLRIMLPAPAGGPSISQGLPASRYNTAADALAAADASAFAOMDVSKTESKHEAA	1847
Dd	2450	-ptlir-rkleeasafels---pssrspasptrsqatpvlspslpdmsls-thssvqagg	2503
Qy	1848	--RLEENLRSRSAAVSEQQOLQOKLTVEKRVSQCCLYTSSAPPCKPQPHSSVVYSAGK	1905
Dd	2504	wrlklppul---sptieyndgrpakrhdiar-----shsespsrlplnrsgtwkrehs	2553
Qy	1906	DKGPPPKSRYEELIRGXTITAAN-----FIDVLIITRIASDKDARERSQSDDS	1957
Dd	2554	hssslp--rvstwrirtgssallsassessekakasedekhvnsisgtkqskengvsakt	2611
Qy	1958	SSSLSSHRYE--TPSDAIEIVSPASSPAPQEKLQTVPEVVKANAENDPTRQVEGPLHH	2016
Dd	2612	wrlkenefaptnstqtvssgatngaektliymapavsk---tedvwriedcpinn	2668
Qy	2017	YRPQESPSPOQLPP-----SSQAEGMGQVPVTHRLITLADHIC	2056
Dd	2669	pr---sgrsptgntppvldsvsekanpnikdkdnqakqnvngsvpm--rtvglenrln	2723
Qy	2057	QIIITQDFARNQVSSQTP-QOPTSTIFQNSPSALVS-TPVRTKTSNRYSPESQAOSVHHQR	2114
Dd	2724	sfiqvdpadckgteikpggmnpvpvsetnessivertpfssssskhspsgstva----	2778
Qy	2115	PGSRVSPENLVDRSGRSRPCGKPSERSHVSEPYEPISPQPVPVWH--EKQDS	2164
Dd	2779	--arvtpfny-----npsrktssadtarspsqip-tpvnmntkkrd	2818
 RESULT 23			
ID	AAW76821	AAW76821 standard; Protein; 2973 AA.	
AC	AAW76821;		
DT	25-JAN-1999	(first entry)	
DE	Human APC protein.		
KW	Tcf; beta-catenin; human; drug; familial adenomatous polyposis; FAP;		
KX	cancer; adenomatous polyposis coli; APC; neoplastic.		
OS	Homo sapiens.		
PN	WO9841631-A2.		
PD	24-SEP-1998.		
Pf	20-MAR-1998; 98MO-US05506.		
PR	20-MAR-1997; 97US-0821355.		
PA	(UYUT-) RIJKSUNIV UTRECHT.		
PA	(UYJO ) UNIV JOHNS HOPKINS.		
PI	Barker N, Clevers H, Kinzler KW, Korinek V, Morin PJ;		
PI	Sparks AB, Vogelstein B;		
DR	WPI; 1998-531569/45.		
XX	Intron-free DNA encoding Tcf-4 protein - useful for,		
PT	identifying drugs for treating FAP patients, or patients with		
PT	increased risk of developing cancer		
PS	Disclosure; Page 32-37; 58pp; English.		







Db 2504 wrklppln1---sptleyndgrprkrdhdiar-----shsespsrlpnrsgtwkrehs 2553  
 QY 1906 DKGPPPKRYEELTRGKTTITAAAN-----FIDVLIIRQIASDKDAREGSGSSDS 1957  
 Db 2554 hasslp--rvstwrtsqssslsasseesekakeedekhnvsisgtkqskengvsakgt 2611  
 QY 1958 SSSLSHRYE-TPSDAIVISPASSPAPQEQKLOTQYQPEVVKANOAEENDPTRQYEGPLHH 2016  
 Db 2612 wrkikenefspntnstqvtvgatngaesktliqmapavsk---tedvwriedcpinn 2668  
 QY 2017 YPQOESPSQOQLPP-----SSQAGMGQVPTTHRLITLIADHIC 2056  
 Db 2669 pr---sgrstgntppvdsysekanpnldskdnqakngvngsvpm--rtvglienrln 2723  
 QY 2057 QIITQDFARNVSSQTP-QQPTSTFQNSPALVS-TPVRTKTSNRYSPESQAOQVHHOR 2114  
 Db 2724 sfliqvdqkgtelkpgnnpvsetnessivertfssssskhspsgtva----- 2778  
 QY 2115 PGRVSPENLVNDRSGRPGKSPERSHVSSEPYEPISPPQVPPVH--BKQDS 2164  
 Db 2779 --arvtpfny-----npsprkssadstsarpsqip-tpvnnntkkrd 2818

RESULT 24  
 AAY70304  
 ID AAY70304 standard; Protein; 2973 AA.  
 XX AAY70304;  
 AC  
 XX  
 DT 06-JUN-2000 (first entry)  
 DE  
 DE  
 KW Human Tcf-4E protein; transcription factor; beta-catenin.  
 KW Adenomatous polyposis coli; APC; transcriptional activation;  
 KW Tcf responsive reporter gene; APC transcription regulatory pathway;  
 KW familial adenomatous polyposis; FAP; cancer; colorectal; thyroid; brain;  
 KW medulloblastoma; breast; head; neck; desmoid tumour; osteoma;  
 KW  
 KW  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200011195-A1.  
 XX  
 PD 02-MAR-2000.  
 XX  
 PF 20-AUG-1999; 99WO-US18774.  
 XX  
 PR 20-AUG-1998; 98US-0136605.  
 XX  
 PA (UYXJ ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX  
 PI He T, Vogelstein B, Kinzler KW;  
 XX  
 XX WPI; 2000-237657/20.  
 DR  
 XX  
 PT Determining wild-type adenomatous polyposis coli protein for diagnosing  
 PT cancer comprises introducing a Tcf responsive reporter gene having  
 PT upstream sequences of c-MYC into a cell -  
 XX  
 XX Disclosure; Page 58-65; 70pp; English.  
 XX  
 CC The present sequence is a protein involved in cancer diagnosis  
 CC associated with APC or beta-catenin mutations. Human Tcf-4E protein  
 CC expressed in colorectal epithelium transactivates transcription  
 CC when associated with beta-catenin. Adenomatous polyposis coli (APC)  
 CC regulates this transcriptional activation, at least in part by binding  
 CC to beta-catenin. Determining wild-type APC protein for diagnosing cancer  
 CC comprises introducing a Tcf responsive reporter gene having upstream  
 CC sequences of c-MYC into a cell and measuring transcription of the  
 CC reporter gene. The candidate drug identified is useful for treating  
 CC familial adenomatous polyposis patients with APC or beta-catenin  
 CC mutations and patients with increased risk of developing cancers such as

CC colorectal, thyroid, brain, medulloblastoma, desmoid tumour, osteoma,  
 CC breast, head and neck.  
 XX  
 SQ Sequence 2973 AA;  
 Query Match 2.9%; Score 365; DB 21; Length 2973;  
 Best Local Similarity 19.0%; Pred. No. 2.9e-11;  
 Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
 QY 347 PEIRKQRQOE-RFORVGQRGAGUSATTARSEH-----EISEIIDLSEENNEK 395  
 Db 981 psiesyeddeskfcysqgypadlahklnsanhmdndgeidtpnysalkysdesqlnsg 1040  
 QY 396 QMRQLSVIPPMFEDAEQRRVKFINNGLMEDPMPYKQDFQFNMVWTDHE--KEIFKDFI 453  
 Db 1041 q-----spsqnerwarpkhileideikqsegrsqrsngstttypvtestcddkhl 1087  
 QY 454 QHPKNFGLIASYLERKSVPCVLVLYLTKNENKALVRRNYGKRGRNQQTARPSOEK 513  
 Db 1088 kfqhfg-----qgecvspy-----rsrgangsetnrvgnhginqvsgqlcqe- 1132  
 QY 514 VEEKEDKAETEK--BEKKDEEKEDEKEDSKENTWEKIDKIDCTAEETEREQATPRG 571  
 Db 1133 -dvyeddkptnyseryseehheerptnysikyn-eeekrhvdpidysikyadipss 1190  
 QY 572 R-----KTANSQGRRKGRITRSMTNEAAS-----AAAAAATPEPPPLPPP 614  
 Db 1191 qkqsfssksssgdsktehmssssentstpsnakrqnqlhpsaqsgrsgqpkaatck 1250  
 QY 615 PEPISTEPVEV-----SRWTEEMEVAKKGLVEHGRN-----WAAIAKM 653  
 Db 1251 vssinqetiqcyvedtpicfsrcsslsssaedeigcncqtqeadsantliqlaeikek 1310  
 QY 654 VGTKE-----AQCNFYFNKRRHNLNLLQHK-----QKTSR 688  
 Db 1311 igtrsaedpvsevpavsqhprtksrllqgsslsasesarhakavefssgskspksagatpk 1370  
 QY 689 KPREE-----RDVSOQSVASTVSAQDEDEIASNEENPEDESEAVEVKPSE--DSPEN 740  
 Db 1371 sppehyvqetplmfsrctsv-ssidsfseisiasvqepcsgmvsgliisdlpdsppq 1429  
 QY 741 A--TSRGNTPEPAVELEPTTETAPSTPSLAVPSTKPAEDESVEVQVNDISIAETAQMDV 798  
 Db 1430 tmppsrsktppp---ppqtaqtkrevpknkaptaeakresgpkqaavnaavq-rvqvlpda 1485  
 QY 799 DQOEHSAEE---GSVC-----DPPPATKADSDVDVEVR---VPEN-HASKVEGDN 840  
 Db 1486 dliihfatestpdgfcscsslsaisldpfigk----dveirimpvqndngneteseq 1541  
 QY 841 TKE--RDLDRASEKVEPRDEDLVVAQIQNAQRPEPQSDNDSSATCSADEVD----- 890  
 Db 1542 pkesnengeakektidsekdl-----dsd-----dddlielleeclis 1581  
 QY 891 ---GEPERQRMFMDSPKSLNPTGSLVSSPLKPNPLD-----LP-----QLQHRAAVIP 938  
 Db 1582 amptkssrkakpqaataasklppvar-----kpsqlpvylkllpsqrllpqkhvstfp 1634  
 QY 939 ---PMWSC---TPCNPIPTGTVSGYAL-----YORHIKAMHES 970  
 Db 1635 gdmprvrycvegtptinfstatslsldtiespnelaagevrggaqsgefekrdtpteg 1694  
 QY 971 AL-----LEQRQREQIDLECRSSSTPCGTSKSPNREWEVL----- 1007  
 Db 1695 rstdeagggktsstvitpeidndkkaeegdilaeacinsampkshkpfrrvkkimdvqqas 1754  
 QY 1008 --QPAPHQLITNLPEGVRLPTTRTPPPPLI-----PS 1039  
 Db 1755 assapkn---nqidgkkkkptsvpkplpqnteytrvrknadsknnlnaervsdnks 1811  
 QY 1040 SKTIVASEKPSF-----IMGSGSIS-----QGTPTGYLTSHNOASTQETPK 1080  
 Db 1812 kkqnknnskdfndklpnndrvrgstafdsphhytlegtp--ycfsrnd----- 1860



Qy 1081 PSVGSISLGLPQOESAKSATLPYIKOBEFSPRSONSQEGLLVRAQHEGVVGRGTAGAIQ 1140  
Db 1861 -slsldfddddvslsrekaelrakenkeseakvtshte--ltsnqgsa--nktqgia 1914  
Qy 1141 EGSITRGPTSKISVESTSPSLRGSTQCTPALPONGIPEALVK-----GS 1186  
Db 1915 kpinrgqpkpilkqs-----tfqsskdipdrgaatdeklqnfalientpvcfshns 1968  
Qy 1187 ISRM-----PIEDSSPEKGRGEEAASKGHVIECKSGHI----- 1219  
Db 1969 isslsldqennnkenepeketpddsgqepskpqasgyapksfhvedtpvcfrnssls 2028  
Qy 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKRYSVEGNIKQGMRESVPVSAPLE 1270  
Db 2029 slsdseddllqecissampkpkksrllkgdnekhspnmggilgeditldkdiqrps 2088  
Qy 1271 GLICRALPRGSHDLKERTVLSSIMOGTTPRATESPEDGLKYPKQIKRESPPIRAFEG 1330  
Db 2089 e---hglspdsenfdkalqegansivsslhqaaaaac-----lsrqassdsdlsilks 2140  
Qy 1331 AITGKPYDGIITIKEMGRSITHEIPRODILLQESRKTEPVQSTRPIIEGSIQGTPIKF 1390  
Db 2141 gislgspf-----hltpdqeekpftknkgprilk-----pgekstletkki 2181  
Qy 1391 DNNS-GQSAIKHNKSLITGP-----SKLSRGM-PPEITVPENIKVVERGYEDVKAGETV 1444  
Db 2182 esekgikgkkykksllitkvgrnseisgmkqplq---anmpsigr----- 2227  
Qy 1445 RSRTSVVSSGSPVLRSLTHLHAPKAQLSPGIYDITSARRTPVSYQNTMSRGSPPMNRITSD 1504  
Db 2228 -----rtmih-----lpgv-----rnsststspvskgqpp 2252  
Qy 1505 VTIPPNKSTNHERKSLTPT-QRESIPAK-SPVPGVDVPVSHSPFDPHHRGSTAGEVWS 1562  
Db 2253 lktpaskspsegqtattspgkpsvkselspvarqtsqigsskapsrgsr----- 2305  
Qy 1563 HLPTQLDPAMPFHRALDPAAYALFQRLSPGVPYQYQLYAMENTRQTILNDIYITSOQ 1622  
Db 2306 -----datpsrpaqpl-----srip 2320  
Qy 1623 MOVNLRPDVA---RGLSPREPGLPLYPATRGIIIDLTWNPTILVPHPGGTSTPPMDRIT 1679  
Db 2321 lqspgrnsispgngisppnk-----lsqilprt-aspstastksgsgkms 2365  
Qy 1680 YIPGTQIFPFRPNYSMSFGPHTHLAAASAEERERERE-----KERERE 1727  
Db 2366 y-----Lspgrmsqnit--kqtglsknassiprsesaskglnqmnngankkvels 2417  
Qy 1728 RIAAASDLXLRPGSEQPGRGSHGYVRSPSPSVRTQETMLQORPSVFQGTNGTSVITPL 1787  
Db 2418 rmssttkss-----gsesq-----rserpvlvrgstfikeaps----- 2449  
Qy 1788 DPTAQLRLMPLPAGPSPISQGLPASRYNTAADALAAALVDAASAPQMDVSTKSKHEAA 1847  
Db 2450 -ptlr-rkleesastesis---psrpsasrtdsqatpvlspisldmsls-thssvqagq 2503  
Qy 1848 --RLEENLRSAVSEQQLEQTELEKRSVOCVLTSSAFPSGKQPHSHSVVYSAGK 1905  
Db 2504 wrklppnl---spcieyngdrpakhdiar-----shsesrplrnsgtwkrehsk 2553  
Qy 1906 DKGPPPKSRYEELTRKTTITAN-----FIDVIITQIASDKDAREGSSQSDS 1957  
Db 2554 hssslp--rvstwtvrtgssslsassesekaksedekhvnsisgktqskencvsaagt 2611  
Qy 1958 SSSLSHRYE-TPSDATEVISPASPAPQKLOTQYQEVVKANQANENDPRQYEGPLHH 2016  
Db 2612 wrklkenefstnstsqtvsngatngaesktliygmapavsk-----tedvvrriedcpinn 2668  
Qy 2017 YRPOQESPSQQLPP-----SSQAEGMGOVPRTHRLITLADHIC 2056  
Db 2669 pr---sgrsptgntppvldsvekanpnikskdnqaknvgngsvpm--rtvglelnrin 2723

Qy 2057 QIIITQDFARNQVSSQTP-QQPSTTFQNSPSALVS-TPVRTKTSNRYSPESQASVHHQR 2114  
Db 2724 sfliqvadpqdkgtelkpgqnpvpvsetnessiverpfssssskhspsgtva----- 2778  
Qy 2115 PGRSVSENLVDRSGRPGKSPERSHVSSEPIPIPPQVPPVH--EKQDS 2164  
Db 2779 --arvtpfny-----npsprkssadstsarpsqip-tpvnnntkkrds 2818  
RESULT 25  
AAV72782  
ID AAV72782 standard; Protein; 2973 AA.  
XX  
XX AAV72782;  
XX  
DT 31-MAY-2001 (first entry)  
XX  
DE Transcriptional activation protein #1 related to the invention.  
XX  
XX Human; Tcf-4 protein; transcriptional activation factor; beta-catenin;  
KW apoptosis; colorectal cancer; Familial Adenomatous Polyposis; FAP; APC;  
KW adenomatous polyposis coli; recombinant adenovirus; Ad-Mini-Me; therapy;  
KW GFP; green fluorescent protein; GFP/cAPC fusion protein; cytostatic.  
XX  
XX Homo sapiens.  
OS  
PN W0200116167-A2.  
XX  
PD 08-MAR-2001.  
XX  
XX 29-AUG-2000; 2000WO-US23635.  
XX  
XX 01-SEP-1999; 99US-0388354.  
XX (UYJO ) UNIV JOHNS HOPKINS.  
XX  
XX Barker N, Clevers JC, Kinzler KW, Korinek V, Morin PJ, Sparks AB;  
PI Vogelstein B, He T;  
XX  
XX WPI; 2001-226675/23.  
XX  
XX New fusion protein comprising an enzyme covalently linked to a portion  
PT of the adenomatous polyposis coli comprising its beta-catenin binding  
PT domain useful for inducing apoptosis or treating colorectal cancer .  
XX  
XX Disclosure; Page 74-81; 83pp; English.  
XX  
XX The present sequence is a transcriptional activation protein related to  
CC the invention.  
CC The invention relates to human Tcf-4 proteins and their corresponding  
CC cDNA molecules which encodes transcriptional activation factors of human  
CC Tcf/Lef family. Human Tcf-4 binds to beta-catenin and activates  
CC transcription in colorectal epithelial cells. Moreover it has been found  
CC that adenomatous polyposis coli (APC) regulates this transcriptional  
CC activation, by binding to beta-catenin. The invention also provides a  
CC recombinant adenovirus, Ad-Mini-Me i.e., APC Minus its amino- and carboxyl  
CC -terminal Ends which expresses a fusion protein, green fluorescent  
CC protein (GFP)/cAPC containing GFP fused to the central third of APC which  
CC contains its beta-catenin binding domain useful for inducing apoptosis or  
CC treating colorectal cancer. These fusion proteins are useful for treating  
CC cancer, e.g. colorectal cancer, and other cancers associated with  
CC Familial Adenomatous Polyposis (FAP) or patients with increased risk of  
CC developing cancer. Human Tcf-4 cDNA provides an excellent system for  
CC screening agents for their ability to promote delivery, integration,  
CC hybridization, expression, replication or integration in cells or in an  
CC animal. It also provides methods for diagnosis cancer in a sample  
CC suspected of being neoplastic.  
SQ Sequence 2973 AA;  
Query Match 2.9%; Score 365; DB 22; Length 2973;  
Best Local Similarity 19.0%; Pred. No. 2.9e-11;











QY 1563 HLPQLDPAHPFHRALDPAAYLQRLQSGTPGYPQSOYQLYAMENTROTILNDYITSQQ 1622  
Db 2306 -----dstpapaqpl-----srip 2320  
QY 1623 MOVNLRPDVA---RGLSPREQPLGLPYPATRYGIIDLINMPPTILVHPGCTSTPPMDRIT 1679  
Db 2321 iqspgrnsipgrngisppak-----lsqiprt-ssptastcksgsgkms 2365  
QY 1680 YIPGQTITFPFPRYNASMSGPHPTHLAAASABERERERE-----KERERE 1727  
Db 2366 y-----tsgrqmsqnlc--kqtglsknassipresasakngnmgngankkvels 2417  
QY 1728 RIAAASDLYLRPGSEOPGRPGSHGYVRSPSPSVRTQETMLQQRPSVFGTNGTSVITPL 1787  
Db 2418 rmsstkss-----gsesd-----rserpvlrvgtflikeaps----- 2449  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALALVDAASAPQMDVSKTKESKHEAA 1847  
Db 2450 -ptlr-tkiesasafesis---psrpsasptrsqaqtpvpslpdmsais-thsvsqaag 2503  
QY 1848 --RLEENLRSRSAVSEQQOLEQKTLVEKRKSVQCLYTSSAFPSGKQPHSSVVYSEAGK 1905  
Db 2504 wrklppnl---sptieyndgrpakrhdia-----shsesparlpinrsgtwkrehs 2553  
QY 1906 DKGPPKSRVEELRKGTTITAN-----FIDVITRQIASDKDARERQSQSDS 1937  
Db 2554 hssslp--rvstwrvtgssslsasessessekaksekdhvnsisgckqscnqvaskgt 2611  
QY 1958 SSSLSHRYE-TPSDALEVISPASSAPPQEKLOTYQEVVKANQANENDPTROYEGPLHH 2016  
Db 2612 wrkikeneftsptnstqtsvsgatngaesktllyqmapavsk-----tedvvrriedcpinn 2668  
QY 2017 YRPOQESPSQOOLPP-----SSQAEGMGQVPRTHRLITLADHIC 2056  
Db 2669 pr----srsptgntppvidsvsekanpnkdkndqakqngvngsvpm--rtvgleln 2723  
QY 2057 QIIQDFARNQVSSQTP-QOPPTSTFQNSPVALVS-TPVRTKTSNRYSPESQAQSVHHQR 2114  
Db 2724 sfqidpddqgkteikpgqnpvpvsetnessivertpfssssskhspsgtva----- 2778  
QY 2115 PGSRVSENLVDKSRGSRPGKSPRSHVSEPEYPISPQVPV-----HEKQDS 2164  
Db 2779 --arvtgfn-----npsprks--sadsys-arpsqiptvnnntkrds 2818  
RESULT 27  
ID AAY57453  
AC AAY57453  
XX AAY57453;  
XX 22-FEB-2000 (first entry)  
XX Human transcriptional regulatory factor SEQ ID NO:10.  
XX DE  
XX KW Human; transcriptional regulatory factor; TCOA1; BLAST detection;  
XX KW bromo-domain; cell proliferation; cancer.  
XX OS Homo sapiens.  
XX PN WO9957143-A1.  
XX 11-NOV-1999.  
XX PF 30-APR-1999; 99WO-JP02340.  
XX PR 30-APR-1998; 98JP-0137631.  
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX PI Jones MH;  
XX WPI; 2000-052940/04.  
XX DR

DR N-PSDB; AA239033.  
XX Transcriptional regulatory factor containing a bromo domain and gene  
PT TCOA1 encoding it  
XX  
XX Claim 1; Page 139-151; 154pp; Japanese.  
XX The present sequence represents a human transcriptional regulatory factor  
CC containing a bromo domain. The factor interacts with proteins involved  
CC in the chromatin-mediated transcription regulation mechanism. It binds  
CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening  
CC compounds binding to it and acting as agonists or antagonists, which  
CC are potentially useful for the treatment and prevention of cancer and  
CC other cell proliferation disorders.  
XX  
SQ Sequence 2781 AA;  
Query Match 2.8%; Score 356.5; DB 21; Length 2781;  
Best Local Similarity 17.6%; Pred. No. 7.7e-11;  
Matches 449; Conservative 354; Mismatches 889; Indels 865; Gaps 108;  
QY 97 SPVDHDSLEKRRPRLEQVSDSHFQVSAAVPLVHPPLPEGLRASADAKKDPAGGKHEAP 156  
Db 270 savyh--lecvkplleevpedewq-----cevcvahkvpvgtdcvaeiqnkpy----- 316  
QY 157 SSPISGQPCG-----DDQNASPSKL-----SKEELIQSDMRVDRDREA 193  
Db 317 ---lrhepigydrsrkywflnrllieedtenenekkiwyystkvqlaelidldkkyw 373  
QY 194 KVEQOILKLKKQOOLEEAAKPEPEKVPSPVPEQKHSRIVQIYYDENRKK-----AE 248  
Db 374 eae-----lckileemreei-----hrhmditedltnkargnsksflaa 412  
QY 249 EAHKIFEL-CPKVELPLYNQPSDTKYVHENIKTNQVWRKKLILFFKRNHARKQREKI 307  
Db 413 aneilesirakkgididnvspeete-----kdknetendskda- 451  
QY 308 CQRYDQLMEAMEKKVDRITENPNRKAESKT-----REYVEKQFPEIRKQREQQERFORVG 363  
Db 452 -----ekneefedqslsksddktpdddpqgsksevgdfksekangelsepg 500  
QY 364 QRGAGLSAT-----TARSEHEISEIIDGLSE--QENNEKQMROL 400  
Db 501 -agkagsgstriitrlrnpdskslsqksqvaahanklfkegkevlvvnsgseisrl 559  
QY 401 SVIPPMFEDAEORRVKFTNMNGL---WEDPMKYVKDQFMNVT-----DHEK-EI 447  
Db 560 stkkevimkg-----ninnfyklgqegkyrvyhnqystnsfalnkhqhredhdkrrh 611  
QY 448 FKDKFIQHPK-NFGLIASYLERKSVPCVLYYLTKNENY-KALVRRNYGKRRGR----- 501  
Db 612 lahfclitpagefkwnsvhgskvltistrltitqlennipssfilhpnwshranwika 671  
QY 502 NQIARPSQ-----EEKVEEKEDKAETKEKE 529  
Db 672 vqmsckprefalalailaecavkpvvmlpiwreflghtrhlrmtsiereekkevkkkq 731  
QY 530 EEKDEEKEDEKESKENTKEKIDGTAETETEERQATPRGRKTANSQGRKGRITRSM 589  
Db 732 eee-----emqqt----- 741  
QY 590 TNEAAAAASAAAAATEPPPPPLPPPPPEIPSTEPVETSQWTE--EEMEYAKKGLVEHGRNW 647  
Db 742 -----wvkytfvkhqwkqgeeyrvtyg-----gsw 771  
QY 648 AA-----IAKVMGKSEAQCKNFYFNKR-----RHNLN-LNLOQHKQKTSRKPRE-- 692  
Db 772 iskthvyrfvplpg-----ntnvnrykslegtknmdenmdesdkrkcspkik 823  
QY 693 -----ERD-----VSOCESVASTSAQEDIED--ASNEENP----- 723  
Db 824 iepdsekdevksdaakgadqnmndiskitekdkdqvdkellidsdskpckkeepmevddm 883















































[illegible][illegible]



QY 482 KKNENYA-----LVRNNGK-----RRGNQOIARPSOEKVEE-----KEEDKAETE 526  
Db 168 kdkneksdvtqlitvkdddegagavagdhqe---psvetavgesakeselkste 224  
QY 527 KKEEKKDEEKEDE---KEDSKENTKEDKIDGTAETETEEQATPRGRKTANSQGRRMG 583  
Db 225 kqegtlkqedssteipiqaesdgaeeekdeg-eeekqeke----- 264  
QY 584 RITRSMYNAAAASAAAAATEEPPLPPPPPISTPEVETSRWTBEEVAKKGLVEH 643  
Db 265 -----ptksp-----espspnsettss-----kffth 290  
QY 644 GRNWAATAKMGVTKEACKNFYNYKRRLNDLLOQHKOT-----SRKPREERDVQ 698  
Db 291 g-wagwrkktfskksdledletaeakrkeqaekvdeeketepaseeqepaedtdqar 348  
QY 699 CESVASTVSAQ-ED-E-DIEASNEE-----NPDESEVA-----VKPSEDSPEN 740  
Db 349 lsadyekvelpledqvgdleaaseekcaplatevfdekmeahqevvaevhstvtekteee 408  
QY 741 ATSRGNTPEAVELEPTTETAPSTSPSLAVPSTKPADESIVETQVNDISIAETAQMDVDQ 800  
Db 409 qggggaeggvvvegtgeslp--peklaepqevpgeapaeelmk-----sremcvsg 459  
QY 801 QESAEEGSGVCDPPATKADSDVDVEVRVPHENHASKVGEDNTER-----D 845  
Db 460 gdht-----qlcdlspek-----tlphpegivsevmssqerikvqgspklkfssg 510  
QY 846 LDRASEKVEP-----RDELDVVAQIINAORPEQSDNDSSATCSADED----- 888  
Db 511 lkklsqkkgkqrgggdeegpgeyghintesadesaqkessasspeeeettclekgp 570  
QY 889 -----VDEPERQRMFMDSPKSLNPNPGSILVSPKPNLDLPQLQHRA 934  
Db 571 leapdgaeeegttsgdckregitpwasfkkmvtpkrv-----rrpsed-----ke 619  
QY 935 AVIPMVSCPTCNPIGTPVSGYALYORHIKAMHESALLEEORORQOIDLCSRSTSPC 994  
Db 620 eelekvksat-----lastdvsemqdevktvgeekpkepkrvd----- 661  
QY 995 GTSKSPNREWEVLAPAPHLITNLPEGVRLPTTRPPPLIPSSKTTVASEKPSPTMG 1054  
Db 662 -tsvs-----weal-----icvgskkarkarkassddg 689  
QY 1055 GSISQGTPTLTSH--NQASYTOETPKPSVGSISLGLPQOESAKSATLPIYIKQEFSP 1112  
Db 690 gprtlgg-----dshraeeaskkeagtdavpast-----qeqdgaqgsssp-----ep 733  
QY 1113 RSQSQPEGLIVRAQHEGVW---RGTAQIQEGSITRGTPTSKISVESIPSLRGSITQGT 1169  
Db 734 agslsegegstwesfirlvtprkkskieeka--edssveqlsteiepsreeswvsik 791  
QY 1170 PALPQTGIPTEALVKGISIRMPIEDSPKREPAASGHVYIEGKSGHILSYDNIK--- 1226  
Db 792 kfipgr---rkradgkqeatvedsgpvneiddpnvpavv-----plseyनावरेक 842  
QY 1227 -NAREGTFSPR-----TAHEISL-----KRSYSEVEGNIKOGMSRESFVS 1266  
Db 843 meagntelpqllgavvyvseelskltvhtsvavldgtravtsve-----erspswisavst 899  
QY 1267 APLGLCLRALPRGSPHDLKERNVLGSGIMQGTTPRATTESFEDGLKYPKQIKRESPIR 1326  
Db 900 eplehtageamp---pveevtekti-----laetpvtlqtpegkdahdmvtse---vd 949  
QY 1327 AFEQAITYKPYDITTIKMGRIHEIPQDILTOBSRKT--PEVQOSTRPIIEGSIQ 1384  
Db 950 fteavtatetsealrt-eevteasgeettmdmsavsgitdspttteeatpvqe--ves 1006  
QY 1385 GTPIKFNNSQSAIKHNKSLITGSKL-----SRGMPLEIVPENIKVVERGK 1434  
Db 1007 gvidteeerqtqallqavadkveesqvpatqvtqrgtskaiekveevsedsevlasek 1066

QY 1435 YEDYKAGETVRSRHTSVVSSG-----PSV-----LRSTLHEAP 1467  
Db 1067 ekdvmpkgpvpqeaagahlaqgsetgqatpeslevpertadvhvatcqviklqlmeqav 1126  
QY 1468 KAQLSPGIYDDTSARRTPVSYQNTMSRGSPMMNTSDVTI--PPNKTINHERKSTLPTQT 1525  
Db 1127 apessetltdsetngstpladsdtad-----gtqgdetidsqskataavrgsqv--te 1178  
QY 1526 RESIPAKSPVGVDPVVSHPSPFPHHRGSGTAGEVYNSHLPTQLDPAMPFHALDPAAAY 1585  
Db 1179 eeaataakeepstlp--nnvpagqeh-geepgr-----dvleptqgeltaaavpv 1225  
QY 1586 LFORQLSPTPCYPSQYQIYAMENTROTILNDYITISQOMQNLNRPDVARGLSPRQPLCLP 1645  
Db 1226 laktev---gqgeev-----dwldgekvk-----eeqevf- 1253  
QY 1646 YPATRGIDITNMPPTILVPHPGGTTTPMDRITYIPCTQITFPFPPRYNSASMSGPHPTH 1705  
Db 1254 -----hsgpnsqkaad-----vtydsevmgvacqcekeste 1284  
QY 1706 LAAAASAEER 1765  
Db 1285 vqslsleegemetdvekeketk-----peqvseege-----qe 1318  
QY 1766 TMLQORPSVFGTNGTSVITPLDPTAQLRIMPLPAGGSPISQGLPASRYNTAADALAAV 1825  
Db 1319 taapeh-----egtygkpvlt-ld-----mpssergkalsg- 1350  
QY 1826 DAAASAPQMDVSKTKESKHEAARLEENLRSAASAVSQOOLEQKTLVEKRKSVOCLYTSS 1885  
Db 1351 -gpslpldq-----kagclevqvgsltdttvtqtaeavekvie----- 1387  
QY 1886 AFPSGKQPQSHSVVYSEAGDKG-----PPKSRVEEELTRGKTTI----- 1927  
Db 1388 -----tvvisetgespecvghllpaekss-----atgghwtlghaedvtplgp 1431  
QY 1928 -TAANFIDVIT-----RQIASDKARERGSQSDSSSSLSHRYE 1967  
Db 1432 esqaespiilvtpeastlhpdlqgeisaqrseeeekpdagpdagdgkesteaiekvlk 1491  
QY 1968 TPSDAIEVSP-----ASSPAPPOEKLQTYQPEVVRANOENDP-----T 2007  
Db 1492 aepeillelesksnkivlnvigtavdgfartetapethaydsqtqvpacridsrepcrwt 1551  
QY 2008 RQYEGPLHHYRQOESPSPQOQLPPSSQAEQMGVPR 2044  
Db 1552 kmkdakmhpvpq-----predlqlvtvleawaq-pr 1582  
  
RESULT 35  
AAW53863  
ID AAW53863 standard; peptide; 1780 AA.  
XX  
AC AAW53863;  
XX  
DT 13-JUL-1998 (first entry)  
XX  
DE Human gravin polypeptide.  
XX  
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;  
KW cAMP-dependent protein kinase; protein kinase C; autoimmune disease;  
KW Myasthenia gravis; nicotinic acetylcholine receptor.  
XX  
OS Homo sapiens.  
XX  
PN US5741890-A.  
XX  
PD 21-APR-1998.  
XX  
PF 19-DEC-1996; 96US-0769309.  
XX  
PR 19-DEC-1996; 96US-0769309.  
XX



(UYOR-) UNIV OREGON HEALTH SCI.

Klauck TM, Nauert JB, Scott JD;

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

WPI; 1998-260552/  
N-PSDB; AAV23545.

New polypeptide fragments of protein kinase binding protein gravin - are useful for the study of modulation of action between gravin and protein kinase(s)

Example 1; Column 19-32; 32pp; English.

This sequence corresponds to the human gravin polypeptide, and represents a polypeptide of the invention. The polypeptides are fragments capable of binding to type II regulatory subunit of cAMP-dependent protein kinase (PKA). Gravin is a kinase anchoring protein that binds to type II regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an antigen of the autoimmune disease Myasthenia gravis (MG), where a patient develops antibodies against their own nicotinic acetylcholine receptors. The polypeptides are useful for providing analogues of gravin in the study of the modulation (e.g. blocking, inhibiting and stimulating) of interactions between gravin and kinase. The peptides are involved in the modulation of gravin-kinase interactions.

Sequence 1780 AA:

Query Match 2.68; Score 323.5; DB 19; Length 1780;

Best Local Similarity	18.6%	Pred. NO. 2.7e-09;	
Matches 361;	Conservative	273;	Mismatches 741;
Indels	565;	Gaps	82;

QY 496 GKRRGNQIARPSQEEKV-----EKEEDKAETEKKEEEKKDEEEKDEKEDSKENT 548

[illegible]

22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53-54-55-56-57-58-59-60-61-62-63-64-65-66-67-68-69-70-71-72-73-74-75-76-77-78-79-80-81-82-83-84-85-86-87-88-89-90-91-92-93-94-95-96-97-98-99-100-101-102-103-104-105-106-107-108-109-110-111-112-113-114-115-116-117-118-119-120-121-122-123-124-125-126-127-128-129-130-131-132-133-134-135-136-137-138-139-140-141-142-143-144-145-146-147-148-149-150-151-152-153-154-155-156-157-158-159-160-161-162-163-164-165-166-167-168-169-170-171-172-173-174-175-176-177-178-179-180-181-182-183-184-185-186-187-188-189-190-191-192-193-194-195-196-197-198-199-200-201-202-203-204-205-206-207-208-209-210-211-212-213-214-215-216-217-218-219-220-221-222-223-224-225-226-227-228-229-230-231-232-233-234-235-236-237-238-239-240-241-242-243-244-245-246-247-248-249-250-251-252-253-254-255-256-257-258-259-260-261-262-263-264-265-266-267-268-269-270-271-272-273-274-275-276-277-278-279-280-281-282-283-284-285-286-287-288-289-290-291-292-293-294-295-296-297-298-299-300-301-302-303-304-305-306-307-308-309-310-311-312-313-314-315-316-317-318-319-320-321-322-323-324-325-326-327-328-329-330-331-332-333-334-335-336-337-338-339-340-341-342-343-344-345-346-347-348-349-350-351-352-353-354-355-356-357-358-359-360-361-362-363-364-365-366-367-368-369-370-371-372-373-374-375-376-377-378-379-380-381-382-383-384-385-386-387-388-389-390-391-392-393-394-395-396-397-398-399-400-401-402-403-404-405-406-407-408-409-410-411-412-413-414-415-416-417-418-419-420-421-422-423-424-425-426-427-428-429-430-431-432-433-434-435-436-437-438-439-440-441-442-443-444-445-446-447-448-449-450-451-452-453-454-455-456-457-458-459-460-461-462-463-464-465-466-467-468-469-470-471-472-473-474-475-476-477-478-479-480-481-482-483-484-485-486-487-488-489-490-491-492-493-494-495-496-497-498-499-500-501-502-503-504-505-506-507-508-509-510-511-512-513-514-515-516-517-518-519-520-521-522-523-524-525-526-527-528-529-530-531-532-533-534-535-536-537-538-539-540-541-542-543-544-545-546-547-548-549-550-551-552-553-554-555-556-557-558-559-560-561-562-563-564-565-566-567-568-569-570-571-572-573-574-575-576-577-578-579-580-581-582-583-584-585-586-587-588-589-590-591-592-593-594-595-596-597-598-599-600-601-602-603-604-605-606-607-608-609-610-611-612-613-614-615-616-617-618-619-620-621-622-623-624-625-626-627-628-629-630-631-632-633-634-635-636-637-638-639-640-641-642-643-644-645-646-647-648-649-650-651-652-653-654-655-656-657-658-659-660-661-662-663-664-665-666-667-668-669-670-671-672-673-674-675-676-677-678-679-680-681-682-683-684-685-686-687-688-689-690-691-692-693-694-695-696-697-698-699-700-701-702-703-704-705-706-707-708-709-710-711-712-713-714-715-716-717-718-719-720-721-722-723-724-725-726-727-728-729-730-731-732-733-734-735-736-737-738-739-740-741-742-743-744-745-746-747-748-749-750-751-752-753-754-755-756-757-758-759-760-761-762-763-764-765-766-767-768-769-770-771-772-773-774-775-776-777-778-779-780-781-782-783-784-785-786-787-788-789-790-791-792-793-794-795-796-797-798-799-800-801-802-803-804-805-806-807-808-809-810-811-812-813-814-815-816-817-818-819-820-821-822-823-824-825-826-827-828-829-830-831-832-833-834-835-836-837-838-839-840-841-842-843-844-845-846-847-848-849-850-851-852-853-854-855-856-857-858-859-860-861-862-863-864-865-866-867-868-869-870-871-872-873-874-875-876-877-878-879-880-881-882-883-884-885-886-887-888-889-890-891-892-893-894-895-896-897-898-899-900-901-902-903-904-905-906-907-908-909-910-911-912-913-914-915-916-917-918-919-920-921-922-923-924-925-926-927-928-929-930-931-932-933-934-935-936-937-938-939-940-941-942-943-944-945-946-947-948-949-950-951-952-953-954-955-956-957-958-959-960-961-962-963-964-965-966-967-968-969-970-971-972-973-974-975-976-977-978-979-980-981-982-983-984-985-986-987-988-989-990-991-992-993-994-995-996-997-998-999-1000-1001-1002-1003-1004-1005-1006-1007-1008-1009-1010-1011-1012-1013-1014-1015-1016-1017-1018-1019-1020-1021-1022-1023-1024-1025-1026-1027-1028-1029-1030-1031-1032-1033-1034-1035-1036-1037-1038-1039-1040-1041-1042-1043-1044-1045-1046-1047-1048-1049-1050-1051-

Qy 549 KEKDIDGTAEETEEREQATPRGRKKTANSQGRRK-----GRITR 587

Db 143 rnieqpsseleeltqpt---esqandlgfkvkfvkfvgfkftvkdkktekpdvtvqlt 199

[illegible][illegible]

Db 200 vkdegegaagdhqdp slgagaaaskesepkqstekpeetlkreqshaeisppaesgq 259

QY 628 WTEE-EMEVAKKGLVEHGRNWAAIAKMVGTKSEAQCCKNFYFNKRRRHNLDNLLQHQHKQT 686

..  
..  
—

..  
—  
—  
..  
..  
..  
—

..  
..  
..  
—  
  
..  
..  
—  
..  
—  
—  
..  
—  
—

c  
c  
t

[illegible]

QY 687 SRKPREERDVSCESVASTVSAQEDEDIASNEEENPEDSEVEAVKPSDSEDPENATSRGN 746

Db 312 frkpke-----deveasekkkeqepekvdtteedgkaevasekltaseqahpqpasa 364

[illegible]

Db 365 heprlsae yekvelpseeqvsqgpseekpaplatevfdekievevhaevhvstve 424

Qy 782 -----TQVND---SISAETAEQMDVDQEH-----SAEGSVCDPPATKAD-SVD 823

[illegible][illegible]

QY 824 VEV--RVPENHASKVEGDNTKER-----DLDRASEKVEP-----RDEDLV 861

Db 485 ekvlskpppegvvsevemlssqermkvqgsplkklftstg1kklsgkkqkgkggdeesg 544



Db 1573 elqtqahvikads-qdaqetekegeepqasqadpitsa----- 1612  
Qy 1950 RGSQSSSSSLSHRYETPSDAIEIVISPASSAPPQKLTQYQPEVVKANQANDPTRQ 2009  
Db 1613 --keesestavqgah-----gdlskdmseas-----ekmttveve-----gstvnd--qg 1653  
Qy 2010 YEGPLHRYRPOQSPSPQOQLPPSSQABGMG--QVPRTHRLITLADHICQIITO----- 2061  
Db 1654 le-----evlpseseegagtksvpeddghallaerleksiivepkekek 1698  
Qy 2062 ----DFARNQVSS-----QTPQOPPTSTFONSFS-----ALVSTPVRTKTS 2098  
Db 1699 gddvdpnqnsaladitdasggltkesptd-----ngpkqekedagevelqegkvhsead 1754  
Qy 2099 NRYSPESQAOVHHQRPGSR 2118  
Db 1755 kaitpqaeeiqkqeresak 1774

RESULT 36  
AAB15380  
ID AAB15380 standard; Protein; 1780 AA.  
XX AC AAB15380;  
DT 26-JAN-2001 (first entry)  
XX DE Human gravin protein sequence.  
XX KW Human; gravin; PKA RII binding site; myasthenia gravis;  
XX KW kinase anchoring protein; CAMP dependent protein kinase.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Binding-site 265..556  
FT /note= "PKC binding site"  
FT Binding-site 1526..1582  
FT /note= "PKA RII binding site"  
FT Region 1537..1563  
FT /note= "PKA anchoring site"  
XX US6090929-A.  
PN 18-JUL-2000.  
XX PD 19-DEC-1997; 97US-0994570.  
XX PF 19-DEC-1996; 96US-0769309.  
XX PR (UYOR-) UNIV OREGON HEALTH SCI.  
XX PA Klauck TM, Scott JD, Nauert JB;  
XX PI WPT; 2000-523763/47.  
XX DR N-PSDB; AAA74903.  
XX PT Novel polynucleotides useful for detecting gravin in patients suffering  
XX PT from Myasthenia gravis encodes CAMP-dependent protein kinase-binding  
XX PT polypeptide and protein kinase C-binding polypeptide of gravin  
XX PS Claim 1; Column 35-45; 34pp; English.  
XX The present sequence is the protein sequence of human gravin. Gravin is  
XX an A-kinase anchoring protein (AKAP) which is involved in the  
XX localisation of CAMP dependent protein kinase A (PKA) via interactions  
XX between the RII binding region and the PKA regulatory subunit RII. Gravin  
XX is also an antigen found in myasthenia gravis sufferers, and it is  
XX thought that antibodies to it may be useful in modulating the binding of  
XX PKA, and thus aid in the treatment of the disease. The gravin coding  
XX sequence was isolated by first screening a human umbilical vein  
XX endothelial cell cDNA library with serum from a myasthenia gravis  
XX patient, and then searching a human heart cDNA library for sequences

CC resembling the isolated sequence. This was done because the first  
CC sequence obtained was shown to be shorter than the full length cDNA.  
XX SQ Sequence 1780 AA;  
Query Match 2.6%; Score 323.5; DB 21; Length 1780;  
Best Local Similarity 18.6%; Pred. No. 2.7e-09;  
Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;  
Qy 496 KRRGRNQIARPSQEEKV-----EKEEDKAEKTEKKEEKKDEEKKDEKESKENT 548  
Db 83 gqkqalngqalnsgeeevivtevgqrdsevdskematskavvhditdggqeen 142  
Qy 549 KKKDKIDGTAEETEERQATPRGKKTANSQGRK-----GRITR 587  
Db 143 rnieqibssesnleeltqpt---esqandigfkvfkfvgfkftvkkdktekptdvqlt 199  
Qy 588 SMTNEAAAAASAA-----AAAATEPPPLPPPPPISTE-----PVETSR 627  
Db 200 vkdegegaagdhqpslgageaasepkskstekpetlkrqshaieispasgq 259  
Qy 628 WTEF-EMEVAKGLVHRGNWAAIAKMWGKSEAQCKNFYFNKRRHNLNLLQOHKQKT 686  
Db 260 aveckegeekqekpskaesptsvtsetgtfkff-----tgwagwrkts 311  
Qy 687 SRKPRERDVSOCEVASTVSAQDEDEIEASNEENEDSEVEAVKPSDESPPENATSRGN 746  
Db 312 frpkpe-----deveasekkkeqekvdeedgkaevaseklitaseqahpgepaesa 364  
Qy 747 TEP-----AVELEPTTETAPSTPSLAVPSTKPAE--DESEV----- 781  
Db 365 heprlsaeyekvelpseeqvgsgpseeekpaplplatevfdekivhqeavvaevhvstve 424  
Qy 782 -----TQVND-----SISATAEQMDVQDEH-----SAEGSVCDPPPATKAD--SVD 823  
Db 425 erteekteveetagsvpaeeelvgmdaepqeaepakelvkiketcvsgedptggadispd 484  
Qy 824 VEV--RVPENHASKVEGDNTKER-----DLDRASEKVEP-----RDSDLV 861  
Db 485 ekvlskppegvvsevmllsqgmkvqgspklkftstgklksgkqkqkrgggdeesg 544  
Qy 862 VAQINAQRPPEQSDNDSSATCSADE-----VDGEPEPRQ 896  
Db 545 ehtqvpadspdsqeeqkgesasspeeeitclekglaevgqdgaeagatsdgekkre 604  
Qy 897 RMFPMDSKPSLLNPTGSLYSSPLKPNPLDLPLQLOHRAAVIPPMVSVCTPCNIPGTGTPVG 956  
Db 605 gvtpwafskkmvtp----- 618  
Qy 957 YALYORHIKAMHESALLEEQROEQIDLECRSSTPCGTSKSPNREWEVLQAPAPHOLIT 1016  
Db 619 ----kkrvirpses-----dkedeldkvsatlsstestaseemqemkgsveepkee--- 667  
Qy 1017 NLPEGVRLPTRRPPPLIPSKTTTVASEKPSFINGGISISQCTPGTYLTSHNOASYTQ 1076  
Db 668 -----pkrvdtsvswaelicvsgsk-----rarrirssde 699  
Qy 1077 ETPKPSVGSISLGLPROQESAKSATLYIKOEESPSRSONSQPEGLLVRAQHEGVVRGTA 1136  
Db 700 eggpkamgg-----dhqkade--agdkdetgtdgilagsqehdpqgagss 741  
Qy 1137 GAIQEGSITRGTPTSKTSVSEIPSLRGSIQTGTPALPQTGPTALVKGSISRRPIEDSS 1196  
Db 742 speqags-----ptegegvstwesfklrvtrkkskleekeedsdiags-----gvehst 792  
Qy 1197 P--EKGREEAASKGHVIEYEGSKGHILSYDNINKNAREGTRSPRTAHEISLKRKSYESVSGNI 1254  
Db 793 pdepepgee-----Swvsikkfipgrkkrip-----dgkqeqapvedag 831  
Qy 1255 KOGMSMRSPVSA--PL-----EGLICRALPRGSPHSDLKERTVLGSGIMOGTTPRAT 1304  
Db 832 ptganedddvpavvpiyseydaverekaeqaqdgaeqpedkaatevskseisqvhmm 891



Qy	1305	TESPEDGLKYPKQIKRESP-----PIRAF--EGAITGKPYDGTITTIKMGRSIHE	1353
Db	892	aaavadttraatlieerspswisasvteplqveaeaalteevleraviaeeptvte	951
Qy	1354	-IP-----RQDILITQSRKTPPEVQOSTRPIEGSISQGTPIKFDNNSQSAIKHN-----	1402
Db	952	plpenreargdtvvsaealtpeavtaa-----etagpligseegtaasaeeetcmvs	1003
Qy	1403	-VKSILITGPKSLRGMPPLEI---VPENIKVVYRGKYEDVKA-GETVRSRHTSVSYSGP-	1456
Db	1004	avsqldtspdttteaatpvgveggvdp-leeqerrttgevlqavaekvkeesqlptggpge	1062
Qy	1457	SVLESTLH---EAPKAQL-SPGIYDDT-----SARRTPVSYQNTMSRGSP-MMNRT	1502
Db	1063	dvlpqvgraaerpeeqaeasglkktedvvlkvdaqaektpeftggkvvggttpefeka	1122
Qy	1503	SDVT--IPPNKSTNHRKSTLTPTQRESIPAKSPVGVDPVVVSHSPDPHHRGSTAGEVY	1560
Db	1123	pqvtesiesseelvttcqaetlagvksqemvmeqaip---pdsvetpdtsetdgtvpvdf	1179
Qy	1561	WSHLPTQLDPAMPFHRA-----LDPA-----AAAYLFQROLSPPTGYPQOYLAMENTRQT	1612
Db	1180	dapgttqkdeivelheenevhlvprvgteaavpaqkrppa---psfv-vfgeetkeqs	1235
Qy	1613	ILNDYITSQOMQVNLRPDVARGLSPR-----EPIGLP-YPATRGIIDLTNMPP	1660
Db	1236	kmedtletdkdevs-etvsilsktegtgeadqyadektdkdvffeglegsld-tgi--	1290
Qy	1661	TILVPHPGGTSMPDMDRIYI---PGQITFPFR---PYNASMSPGPIHTLAAAASA	1712
Db	1291	-----tvsrkercvvalgkegtgeaeackdddalelqshakspvperemvqv	1339
Qy	1713	ERER-ERE-----REKERERITAAASDLYLRPGSEQ-----PGRPGSHGYVRSP	1757
Db	1340	erekteaethvneeklehetavtveev-----skqltvtvnpildgakevslegsp	1394
Qy	1758	SPSVTRQETM-----LQRPSPVFOGTNGTSTVITPLDPTAQLRIMPL-----PAGGPSISQ	1807
Db	1395	ppclgqeeavctkigvqsseasftlaaaeeckvlgeta--niletgetlepagahlvle	1452
Qy	1808	GLPASRYNTAA-----DALAALVDAASAAPQMDVS-----TKESKHEAAR	1848
Db	1453	ekseknedfaahpgedavptgtpdcakstpvivsattkglssdlegektslkwsde	1512
Qy	1849	LEENLRSAAVS---EQOOLQKLTLEYEKRS-----VOCLYTSSAPFSG	1890
Db	1513	vdeqvacqevkvsvaldilepngillektsklvqnliqtavdqvfvrteetatmlts	1572
Qy	1891	KPQPHSVVYSEAGDKGPPKPSRYEE-ELTRGKTTITTAANFIDVITRIQIASDKDARE	1949
Db	1573	elqtqahvikads-qdagqetekegeepqasaqdetpitsa-----	1612
Qy	1950	RGSQSDSSSSLSHRYETPDAIEVISPASPAPQPEKLOTYOPEVVVKANQAENDPTRQ	2009
Db	1613	--keesestavgqah-----sdiskdmseas-----ektmteve-----gstvnd--qq	1653
Qy	2010	VEGPLHHYRPQQSPSPQOQLPSSQAQMGV-QVPRTHRLITLADHCILCIITQ-----	2061
Db	1654	le-----evlpsseeeggagtksvpeddghallaerleksivpepkedex	1698
Qy	2062	----DFARNQVSS-----QTQQPPTSTFNQSPS-----ALVSPVPRTKTS	2098
Db	1699	gdvdpdenqnsaladtasggltkesptd---ngpkakekedaqevlqegkvhsesd	1754
Qy	2099	NRVSPESQAQSVHHQRPGSR	2118
Db	1755	kaftpqaqeelqkqeresak	1774

RESULT 37  
AAV85657  
ID AAY85657 standard: Protein: 1341 AA.







PT chromosome 11

XX Disclosure; Page 29-50; 90pp; English.

XX This sequence is encoded by the acute lymphoblastic leukemia (ALL-1) gene of chromosome 11. The ALL-1 gene was isolated by translocation breakpoint mapping. Fragments of the ALL-1 cDNA may be used to identify chromosomal abnormalities within the ALL-1 gene. These fragments may be used in the treatment and diagnosis of human leukemias such as acute lymphocytic, myelomonocytic, monocytic and myelogenous leukemia. ALL-1 protein shows three regions of homology to the prosopha trithorax protein. These regions show 64%, 66% and 82% similarity respectively, to the Drosophila gene. The third region of homology constitutes the extreme C-terminus of the two proteins, both proteins end in an identical sequence. The first homology region is cysteine-rich and contains sequence motifs analogous to four zinc finger domains (3-6) within the trithorax gene. The second region of homology is also cysteine-rich and corresponds to zinc fingers 7 and 8 of the Drosophila gene. The multiple conserved cysteines and histidines at the 3' end of the motifs allow two or three arrangements of the putative fingers. The structure of these cysteine-rich domains appears to be unique to the trithorax and ALL-1 genes.

XX Sequence 3910 AA;

SQ Query Match 2.5%; Score 316; DB 14; Length 3910;

Best Local Similarity 17.7%; Pred. No. 2e-08;

Matches 581; Conservative 367; Mismatches 1067; Indels 1264; Gaps: 153;

QY 13 FSTEQSRYPHVSQ--YTFPNTRHQEFVAVDYRSHLEVSQALLQOQQOQLR-RRP 69

DB 366 fiededyppikarlestpnr----fsapcysse-kssaaqshssqmsdssrrssp 420

QY 70 SL-LSEFHPGSDRQ---ERTSYEPFHPGSPVD---HDSLESKPRLEQVSDSHFQRV 122

DB 421 svdstdsqaseeiqlpeicrstdpvhv-plpispgenesndrrrsyvsersfgr 479

QY 123 SAAVLPLVHPLEGLRASADAKDPAGFKHEAPSPISQPCQGDQNASPSKLSRELI 182

DB 480 tkklstlqapq--qetssppppll-----tpppl--qpassisdhtpwm----- 524

QY 183 QSMRDVRETAKEVQOILKLLKKQOQLLEEAARKPP-----EPEKPSPPVPEOKHSIV- 236

DB 525 -----pptiplasplpastapmqqkrksilr 551

QY 237 -----QIITYDENRKKAEAA---HKIFELGPKVELPLYNQPSDF----- 272

DB 552 eptfrwtlksrseppqfssakayakeglirkpifdnfrpp---pl--tpedvgfasgfs 606

QY 273 -----KVYHENIKTNQVMRKKLI---LFFXRNRHARKQREOKICORYDQLMQAW 318

DB 607 asgtaasarlfsphsgtrfdmhkrspllraprftpseahsr-----ifesv 653

QY 319 EKKVDIENPRKAKESKTREYERQFPIRKQREOQERFORVGORGAGLSATIAHSEH 378

DB 654 tlpsnrtsgatsssgsvnrkr--rkvfpisrsepspsmsmt--rsgrlsselspl 709

QY 379 EISEIIDGLS-----EQENNEKQMRQLSVIPPMFDAE 411

DB 710 ppsvasssisvsplatsalnptftfpshltqsgesaekndprkqtsa-paepfss 768

QY 412 QRRVKF-----INMNGLMED--PMKVYKDRQFMVMTDHEIKFQKDF----- 452

DB 769 sptplfpwftpgsqtergnrkkapeelskdr-----daksvekdksrdrereke 821

QY 453 -----TQHPKNFLGILASYLERSKVSPPDCVLYYITKNE----- 485

DB 822 nkresrkrkkgseiqsssalypvgvrskvkvvgvedatssakkatgkkskshdsdt 881

QY 486 -----NYKALVRNRYGKRRNQIARPSOEKVEEK----- 517

DB 882 ditsvtlgttavtktkillkkgrgnlektndlg-ptapslekektlclstpsstvtkhs 940

QY 518 -----EEDKAETKEK-----EEKDEEKEDEKESKENTKEKDIDGTAETEER 564

DB 941 tssigsmiaqadklpmdtkrvasllkkaklaqlckiekskllqt-dqpkaqqgsdset 999

QY 565 EQATPR---GRKTANSQGRKG-----RTRSMTNEAASAAAAA 602

DB 1000 svrgprlikhvcrrraavalkrkvafpdmptlalsalpweereekilssngndkkslsgs-- 1057

QY 603 ATEPPPPPLPPPEPTISTEPVTSRTEEMEVAKGLVEHGRNWAALAKMVGTFKSAQC 662

DB 1058 ---edaelpappikpi--kpvtnkapqpep-----vkkgrrrrcgcgcgcqcpvpedc 1105

QY 663 -----KNFYFNKRRHNDNL-----LQOHKQKTSRKR 691

DB 1106 gvtncldkpkfggrnikkccckmrkcqnlqmpskaylqkqakavkkkkekktsek-k 1164

QY 692 EERDVSOCEVAS-----TVSAQDEDEIEASNE-----EENPEDESEAVKVPSESDSP 738

DB 1165 dskessvknvndvssqktpksaredpapkkssepprpkpveekseegnvsapgeskqa 1224

QY 739 ENATSRGN---TEPAVELEPTTETAPSTSPSLAVPSTKPAEDESVEVQVNDISISAETAE 794

DB 1225 ttpasrksskqvsgpalvippqpt--tgpprkevptktbsepkkq----- 1269

QY 795 QMDVDQOQESAERGSCVCDPPPKADSDVDVEVRVPRNHASKVGEKDNTKRDLDRASEKVE 854

DB 1270 -----pppsesg-----pegskqk-----kva 1286

QY 855 PRDEDLVAAQIINARPEPOSDNDSSATCSADEVDV-DGEPERQRM-----FPMD 902

DB 1287 prpsipvkqpkpkekekepppvknkgenagtlilstlssngsskqkkipadgvrhrizvdfkd 1346

QY 903 SKPSLNLPTG--STLVSSPLKPNPLDLPLQLOHRAAAVPPMVSCPTPCNIPICGTPVSGVALY 960

DB 1347 ceenwemvggliltvvp-----itprvvcflc-----assgvnef 1383

QY 961 ---QRHKAMHESALLEEQROREQID-----LE---CRSST 991

DB 1384 vycqvcepfhkcleenerpleqlecnccrrckfchvcgrhqatqkllecnkcnusy 1443

QY 992 SP-----CGTSKSPNREWEV----- 1006

DB 1444 hpeclgpnypkptkktkvwlctkcrcvckscg-sttpgkgwdaqwhdalfchdcaklfa 1502

QY 1007 -----LQAPAPHOLITNLPEGVRLPT 1026

DB 1503 kgmfcplcdkcydddyeskmmgcgkcdrrwvshkcenlsgtedemyeillsnlpesvaytc 1562

QY 1027 TRPTRPPPLIPSKTTVASEKPSFIMGSIISQGTPTGYLTSHNQASVYQETPKPSVSGSI 1086

DB 1563 vncterhp-----aewrlalekelqisikqvltaallnsrttsh-llryrqaakppdin-- 1614

QY 1087 SLGLPROQESAKSATLP-----YIKQEEESP- 1112

DB 1615 ---peteesiprspegpdpvltvevskgdqddqldlegvkrkmdqgnytvslefsdd 1670

QY 1113 -----RSQNSQPE---GLLVRA---QHEGV-----VRGTAGAI 1139

DB 1671 ivkligaaainsdggqpeikkansmvksffirmervfpwfsvkksrfwepknkvssngml 1730

QY 1140 -----QEGSITRGTPT-TSKISVESIPSLRGSIPTGTPALPOTGIPT 1179

DB 1731 pnavlppslldhnyaqwqereenshteqplmkkilipakpkpgpdpdpplph-----pt 1786

QY 1180 EALVKGSISRMPEDSDSPEKREE-----AASKGHVIYEK----- 1215

DB 1787 ppllstdrsredspelnpppgiednrqcalcltygddsandagrillyiqdnwchvnal 1846

QY 1216 -SGHILSYD--NIKNAK----- 1229

DB 1847 wsaeivfedddgsiknvhmavirqlrcefcqkpgatvgccitcstctsnhyfmsrakncv 1906



























PT - which inhibits and/or promotes the activation of cAMP and mitogen  
 responsive genes

XX Claim 1; Page 32-39; 46pp; English.

CC Identification of a cpd. (1) which inhibits and/or promotes  
 activation of cAMP and mitogen responsive genes comprises monitoring  
 expression of a reporter in response to (1), relative to expression  
 of reporter in the absence of (1), where exposure of (1) is  
 performed in the presence of (1) a signal dependent transcription;  
 CC (2) a polypeptide comprising at least amino acid residues 461-661 of  
 this sequence (AAR79054); and (3) a reporter construct comprising a  
 CC reporter gene under the control of a signal dependent transcription  
 factor. The method can be used to identify a compound which has  
 CC the binding and/or activation properties of CREB binding protein or  
 CC the transcription activation characteristic of a signal dependent  
 transcription factor.

XX Sequence 2441 AA;

Query Match 2.48; Score 309.5; DB 16; Length 2441;  
 Best Local Similarity 18.78; Pred. No. 2.4e-08;  
 Matches 495; Conservative 304; Mismatches 992; Indels 859; Gaps 122;

Qy 7 PPNQAFSTEQSRYPHSHVQVTFNTRHQEFVAVDYRSHLEVSVQASQLLQOQOQOQ-- 64  
 Db 121 plngdst-----pnlpka-----astsgptppasqalnpaqkqvg 159  
 Qy 65 LRRRPSLLSEPHSGDRPQERTSYEPHPG--PSPVDHDSLEKRRPLEQVSDSHF--- 119  
 Db 160 lvtspatsqtgpgic-----mnanfnqthpgllnnsnshslmnaqagqagqvmngslqaa 215  
 Qy 120 QRVSAVLPLVHPLPEGLRASADAKK----DPAPGCKHEA----- 155  
 Db 216 grgrgagmpypapamqagatvslaeltlqvspqmag-haglnataqaggmktmgmtgttsp 274  
 Qy 156 ---PSSPTSGQPCGDDQNASPKLSKEILQSDMRDVRDEIAKVEQIILKLRKQOQLEEE 212  
 Db 275 fgqfsgtggqmg-atgvnqlaskqsmvnslpaftdiknts---vttvpnmqlqts 330  
 Qy 213 AAKPEPEKVPSPVVEQKHSIVQ-----IYDENRKAEEAHKIFELGLPKVELPLYNQ 268  
 Db 331 vgiyptqaiatgptadpekrklqqlvlllhakcqrqean----gevracslp----- 382  
 Qy 269 PSDTKVYHENIKTNQVMKKLILFFKRNHARKOREQKICO-----RYDQLMEAWEK--K 321  
 Db 383 -----hcrf--mknvl-----nmthcqapkcacqvahcaserqlishwkncr 423  
 Qy 322 VDIENNPRRKAKESKREYIEYKOFPEIRKOREQOERF-----QRVGORGAGLSATI 373  
 Db 424 hdcpcvlpknasd-----krnqqtllgspasglqntlgsvgagqgnat 467  
 Qy 374 ARSEHETSEIIDGLSEQF-----NNEKOMRQLSVIP---PMFDEARRVKFINNM 421  
 Db 468 slsn--pnpidpsmggrayaalpymnqptqlqpqvpgqgpaqpbahq---qmrtln 521  
 Qy 422 GLMEDPMKV-----YKQDFMNVTDHEKEI---FKDKFIQHPKNFGLIASYLERKSVDP 473  
 Db 522 algnpmvsvpaggittddqgnplnlsaelptslgatpmlmdgnsngnlsi---stip- 577  
 Qy 474 CVLYYILTNNENTKALVRNRYKRRGNQOIARPSQOEKVEEKEE-----DKAEKTEK 527  
 Db 578 -----taappstgvrkgwheivtdqlshlvhkvqalfptdpaaakdr 623  
 Qy 528 KEEE-----KKDEEKDEKESKNTKE--KDKIDGTAETFEEREQATPRGRKTANSQGR 580  
 Db 624 rmenlvayakkvegdmyesansrdeyhlhlaekiykikqlelekr-----rtrlhqg- 676  
 Qy 581 RKGRITRMTNEAASAAAAATEEPP-----PPLPPPPPEPISTPEVETSRWTEEEVEV 635  
 Db 677 -----ilgnqpalpasgqppvipagvrrpbnqplpvpnmqvsqgmnsfnpmsl 728

Qy 636 AKGLVEHGRNWAIAKM---VGTSKSAQCKNFYFNVKRRHNLNLLQOQKOK-TSRKPR 691  
 Db 729 gnvqlpqapmgpraaspmnhsvqmnsmasvpgmaismrmpqppnmmtgthannimaqpt 788  
 Qy 692 EERDVSOCE-----SVASTVSAQEDIEDIASNEEE-----NPEDESEAV 731  
 Db 789 qnqflpqnqfsssgamsvnsgmgqpaagagvsqgqgaalpnlmlapqasqipc- 847  
 Qy 732 KPSEDSP-----ENATSRGNT--EPAVELEPTEPTAPSTSPSLAVPST 772  
 Db 848 ppvtqslphtpppastaagmpsiqlhtpagnmtppqpaaptqstpvssggtpt-ptpgs 906  
 Qy 773 KPADESEVETQNDISIAETAQMDVQOQHSASEGSCVCDP-----PPAT-- 817  
 Db 907 vps---aaqtqstvtvaaaqav--tpqptqtpvppsvatpqssgqgqptvhtqpgtpl 962  
 Qy 818 --KADSVDEVVRPENHASKVGEQNTKERDLDRASEKVEPRDEDLVAAQINAOQPEPOS 875  
 Db 963 sqaasidnrvtpstvtstaetsqgppdvpmlemktevtqdd-----aepepte 1013  
 Qy 876 DNDSSATCSADEVDGEPERQRMFPMDSKPSLLNPTGSLVSSPLKPNPLDLPOLOHRAA 935  
 Db 1014 skgeprsemmeedlqgssqvkeetd-----tteqksepmeveekkpvk 1057  
 Qy 936 VIPPMVSTPCNIPITGPVSGYALYQRIKAMHESALLEEQORQEQIDLECRS-STSPC 994  
 Db 1058 v-----eakeeeensndtasqstps 1079  
 Qy 995 GTSK---SPNREWEVLOPAPHQLITNLPVGVRLTPTRTPPPPLIPSSKTTVASEKPSF 1051  
 Db 1080 qprkikfpeelrqlmptlealyrqdps--lprfqvdpqligdyfdvknpmlds 1137  
 Qy 1052 IMGGSISOG---TPGTY--LTSHQASVTOETPK-----PSVCSI 1086  
 Db 1138 tikrkidtgqvepqvydvrlmfnawlnrktsvyfkscklaevfegeidpvmqsi 1197  
 Qy 1087 SLGLPROQESAKSATLPYIKOEESPRSQNSQPEGLIVRAQHEGVVVRGTA-GAIEQSIT 1145  
 Db 1198 gycgrkyefspqtlccyqkqlctipr-----daayysyqnrhyfcgkctelqgenvt 1251  
 Qy 1146 RCTPTSKISVESIPSLRGSITQGTIPALPOTGIPTEALVKGISIRMPIDEDSSPEKGREAA 1205  
 Db 1252 lg-----ddpsqpttiskdqfekkndtl---dpepfvdckecg 1288  
 Qy 1206 SKGHVIVEGSKHILSYDNI-----KNAREGTRSPRTAHEISLKRSEYSEVG- 1253  
 Db 1289 rkmhqic-----vlhdyliwpsgfvcdnclktgrprkenkfakrlqtrlnghledr 1342  
 Qy 1254 IKQMSMRSPVSPAPLEGLICRALPRGSPHSD-----LKERTVLGSGTMOQTTPRATT 1305  
 Db 1343 vnkflrrqnhpcaq-----evfvrivassdktevkvpgmksrfdsgemesefpyrkt 1395  
 Qy 1306 ESFEDGLKYPKQIKRESPPITRAFEGATKGPYDGITIKEMGRSIIHE---IPQDILT 1361  
 Db 1396 alf-----afe-----eidgv-dvcffgmhvqdtallaphq----- 1425  
 Qy 1362 QESRKTPEVWQSTRPIIEGS--ISOGTPIKFDN-NSGOSAIKHN-----VKSLLI--- 1407  
 Db 1426 -----lqgcviysilsihfrprclrtavvheillgylevaykklvyvt 1469  
 Qy 1408 -----TGPSKLSRGMPPLEIVPENIKVVERGK-----YEDV--KAGE 1442  
 Db 1470 ahiwacppsegddyfchcpdpdqkpkprkqlewykkmldkafaerlindykdifkane 1529  
 Qy 1443 TVRSRHTSVV-----SSGPSVLRLTHLEAPKAQLSPGIYDDTSARFTPVSYQ-----NT 1491  
 Db 1530 ---drlltsakelpyegdfwvnpvleesikeleqegeeeerkkeestaasetpsgsgdskna 1586  
 Qy 1492 MSRGSPMMNRT-SDVT-----lppnkstnherkstltpores-----IPA 1531  
 Db 1587 kknknktnknksisranksnkmpnvnndlsqklyatmekhkevfvfvlhlagpvlst 1646  
 Qy 1532 KSPVPVGDVPPVVS-----HSPF-----DPHHRGSTAGEVYWS----- 1562















Qy 1361 TQESKTEPVVQSTRPIIEGSIQCTPIKFDNNSQSAIKHNKSLIRGPKSLSRGMPL 1420  
Db 930 psgfpap-tvestpdcvcpaspalqp---slssgq-----lpl 965  
Qy 1421 EIVPENIKVREGKYEDKAGETVRSRHTSVSSGSPSVLRSTLHEAPKAQKSPGIYDDTS 1480  
Db 966 -liptd-----pspppcppvltvatpppllpvtvplpapssasp----- 1005  
Qy 1481 ARRPVSVQNTMSRSPMMNFTSDVTTPPNKSTNHERKSTLTPTQRESIPAKSPVGVDP 1540  
Db 1006 -hpcp-----splsnatagapip-----ilspvts---papsipvpv 1040  
Qy 1541 VVS-HSPDPHRRGSTAGEVWHLPTQLDPAMPFHRALDPAAYALFQRLSPTPGVPS 1599  
Db 1041 lmsaaspplslsssssssssssssssp-----sppp----- 1078  
Qy 1600 QYQYAMENTROTILNDYITSQQOVNLRPDVA--RGLSPREQLPL----- 1644  
Db 1079 ---lsaisvsvs--gdnleasplmiskgeeleneglkpreepqsaeeqvvvqetfkn 1133  
Qy 1645 -----PPATRGII-----DLTNMPTTILVPHGCT--- 1670  
Db 1134 nfvencvcsplfsikdtkhlsiaewpfkcefcvqlfkdtkdlshehrfilhgvgnfv 1193  
Qy 1671 -----STPPMDRI---TYIGTQITPPPPYNSASMSGHPTHLA 1707  
Db 1194 cvckkefafilnqghrdlhp-dkctthefesgtl-----fpqfndpskahvehmq 1247  
Qy 1708 AASAER 1767  
Db 1248 slp-----edpletskeeeelndseelyttikimsg-----lktkdpvrlg--- 1291  
Qy 1768 LQQ-----RPSVFQ-----GTNGTSVITPLDPTAQLRIMPLPAGGPIISQGLP 1810  
Db 1292 lnhypsfkpppfgyhrrnmpmgigvgtatnftthnlpqgtfaiactckgk---vdmnp 1347  
Qy 1811 ASRYNTAADALAAALVDAASAPO---MDVSKTKESKH----- 1844  
Db 1348 elkhkllacasa--dkkryptknpvplkqtpkngvvlndsgknafrmgqpkrln 1405  
Qy 1845 ---BAARLEENLRGRSAVSSQOOLEKTLERKRSVO-----CLYTSSA 1886  
Db 1406 fsvelskmsn--kiklnalkknqlvqkailqknksakqkdlknacessshicpynre 1464  
Qy 1887 FP-SGKQPQSHSVVYSEAGDKGPPKSRYEELTRGKTTITANFDVITRQIASDK 1945  
Db 1465 ftylgslnkhaa--fscpkplspk-----kvs 1492  
Qy 1946 DARBERGQSDSSS---SLSHRYETPSDAIEVISPASSPAPQEKLTQYQPEVVKANQA 2002  
Db 1493 hsskkgghspassadknsnshrrt-----adaeikngsmatplgktrar 1538  
Qy 2003 ENDTRQYEGPLHYRQOE---SPSQQLPPSSQAGMGQVPRTHRLITLADHICQII 2059  
Db 1539 ssgpt-qvpipssfrskqnkvfaasvkskpsssslrnsapi-----rmakit-hvegkk 1592  
Qy 2060 TODFARN---QVSSOTQPPPTSFQSPSAL--VSTPVRTKTSNRYSPESQAS 2109  
Db 1593 pkavahnhsaqlsaktshlrv-qkskavlgsktskrtdrfnikrsers 1646

RESULT 45  
AAW40058  
ID AAW40058 standard; Protein; 2441 AA.  
XX  
AC AAW40058;  
XX  
DT 20-JUL-1998 (first entry)  
XX  
DE Cellular transcriptional factor CBP.  
XX  
KW Cellular transcriptional factor; CBP; P/CAF; human; transcription;

histone acetyltransferase; HIV; infection; cancer; therapy;  
muscle differentiation.  
Homo sapiens.  
Key Location/Qualifiers  
Misc-difference 377 /note= "encoded by CGN"  
XX W09803652-A2.  
XX 29-JAN-1998.  
XX 23-JUL-1997; 97WO-US12877.  
XX 23-JUL-1996; 96US-0022273.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX Howard BH, Nakatani Y;  
XX WPI; 1998-120777/11.  
XX N-PSDB; AAV10093.  
XX New isolated p300/CBP-associated factor, P/CAF - used to develop  
products for modulating transcription, e.g. for treating HIV  
infection or cancers or for promoting muscle differentiation  
Disclosure; Page 81-86; 107pp; English.  
XX This polypeptide sequence comprises CBP, a global transcriptional  
coactivator that is involved in the regulation of various  
DNA-binding transcriptional factors. The invention relates to a  
novel human p300/CBP associated cofactor, P/CAF (see AAW40052), that  
modulates transcription through binding to p300 and CBP (see  
AAW40058). The region (see AAW40060) of CBP that binds to P/CAF  
is claimed. The invention provides methods of screening for  
compounds that inhibit or stimulate the transcription modulating  
and histone acetyltransferase activity of P/CAF and p300/CBP.  
XX Inhibitors can be used e.g. to inhibit HIV TAT-mediated  
transcription in the treatment of HIV infection. Stimulators can  
be used e.g. to activate tumour suppressor p53 in the treatment of  
cancer or to activate the muscle differentiation factor MyoD to  
promote muscle differentiation. The products can also be used to  
inhibit the cell cycle progression inducing effect of an  
oncoprotein which binds p300/CBP in a subject.  
XX Sequence 2441 AA;  
SQ  
Query Match 2.4%; Score 308.5; DB 19; Length 2441;  
Best Local Similarity 18.5%; Pred. No. 2.7e-08;  
Matches 489; Conservative 305; Mismatches 1005; Indels 843; Gaps 116;  
Qy 7 PPNGAFSTEQSRYPHVSQYTFPFRHQEFVAVPDYRSSHLEVSQSLIQQQQQQ-- 64  
Db 121 plhgdsst-----pnlpkqa-----astsgtpptasqalnpqaqkvq 159  
Qy 65 LRRRPSLLSEFHGSDRPPQRRRSYEFPHG--PSPVDHDSLESKRPRLEQVSDSHF--- 119  
Db 160 lvtsspsatsqtgpgic-----mnanfnqthpgllnsnsgshlmnqagqgqagvmngsgaa 215  
Qy 120 QRVSAAVLPLVHPLPEGLRASADAKK---DPAFGGKHEA----- 155  
Db 216 grgrgagmpypapamqgatssvlaetltqvspmag-haglnaqagmtkmgtgtsp 274  
Qy 156 ---PSSPISQPCGDDQNASPSKLSKEELIQSMDRVDREIAKVEQQILKLLKKQKQLEEE 212  
Db 275 fggpfsqtgggqmg-atgvnpqlaskqsmvnsipaftdiknts---vttvpmnsqqlts 330  
Qy 213 AAKPPEPEKVPSPPPVEQKHSRIVQ----IYDNRKKAEEAHKIFEGLGKVELPLYNQ 268  
Db 331 vgiuptqaiatgptadpekrklqqqlvlllhahkcdrrrega----- 372







```

Db 2262 qmaapmgqlgmgqpglqadstpnliqqalqqrilqqqmkqigspgqpnmpspqqhmls 2321
QY 2041 QVPRTRHLITLADHI--CQITQDFARNQVSSQTPQPPPTSTFONSPSALYSTPVRTKTS 2098
Db 2322 gqpq-----ashlpqqiats--lsnqvrsapvqgsprpq-----sqpphssps 2363
QY 2099 NRYSPESOAQSVHH-----ORPGSRVSPENLVDKSRGSRPGKS-----PERSHVS 2143
Db 2364 priqpq---psphhvspqtgtphpglavtmasmdqghlgnpedsamlpqintpnrsals 2420
QY 2144 SE 2145
Db 2421 se 2422

```

Search completed: September 8, 2001, 14:42:18  
 Job time: 480 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 8, 2001, 14:35:13 ; Search time 23.22 Seconds  
(without alignments)  
2163.671 Million cell updates/sec

Title: US-09-522-753-11  
Perfect score: 12643  
Sequence: 1 MSSSGYPNQGAFTSQSRY.....EREPAPLLSAQYETLSDSD 2440

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2885.5	22.8	619	2	US-08-372-652-5
2	2885.5	22.8	619	5	PCT-US95-16311-5
3	368	2.9	2843	1	US-07-741-940-2
4	368	2.9	2843	1	US-08-289-548A-2
5	368	2.9	2843	1	US-08-452-654-2
6	368	2.9	2843	2	US-08-370-235A-2
7	365	2.9	2842	1	US-07-741-940-7
8	365	2.9	2842	1	US-08-289-548A-7
9	365	2.9	2842	1	US-08-452-654-7
10	365	2.9	2843	1	US-08-452-655B-2
11	365	2.9	2843	1	US-08-452-655B-7
12	365	2.9	2843	3	US-08-450-582-2
13	365	2.9	2843	3	US-08-450-582-7
14	365	2.9	2973	2	US-08-821-355A-7
15	365	2.9	2973	2	US-09-003-687A-7
16	365	2.9	2973	4	US-09-136-605-7
17	329	2.6	3969	4	US-08-061-376-5
18	323.5	2.6	1780	1	US-08-769-309A-5
19	323.5	2.6	1780	3	US-08-994-570-5
20	309.5	2.4	2441	1	US-08-194-468-2
21	309.5	2.4	2441	3	US-08-961-739-2
22	309	2.4	1719	2	US-08-459-568-4
23	309	2.4	1719	3	US-08-399-411-4
24	309	2.4	1719	3	US-08-516-859A-4
25	294.5	2.3	2035	1	US-08-046-585-5
26	294.5	2.3	2035	1	US-08-393-703-5
27	294.5	2.3	2035	5	PCT-US93-11721-5

Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 94, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Patent No. 5180808  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 23, Appli  
Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-372-652-5  
; Sequence 5, Application US/08372652  
; Patent No. 5932699  
; GENERAL INFORMATION:  
; APPLICANT: Moore, David  
; APPLICANT: Seol, Wongi  
; APPLICANT: Choi, Hwang-Sik  
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street, Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 619 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-372-652-5

filed 11/3/95 102e

Query Match 22.8%; Score 2885.5; DB 2; Length 619;  
Best Local Similarity 90.1%; Pred. No. 2.4e-171;  
Matches 562; Conservative 27; Mismatches 28; Indels 7; Gaps 3;  
QY 1819 DATAALYDAASAPQMDVSKTKESKHAARLEENLRSAVSEQQQLKLEVKRSV 1878  
|||||



Db 1 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERKSV 60  
QY 1879 QCLYTSSAFPSGKQPQSSVYVYSEAGDKGPPPKSRVYEEELTRGKTTITAANFIDVIT 1938  
Db 61 QCVCTSSALPSGKAQPHASVYVYSEAGDKGPPPKSRVYEEELTRGKTTITAANFIDVIT 120  
QY 1939 QRIASDKDAREGQSQSSSSLSHRYETPSDAIEVISPASSPAPPOEKLTQYQPEVVK 1998  
Db 121 QRIASDKDAREGQSQSSSSLSHRYETASDAIEVISPASSPAPPOEKQYQPDVVK 180  
QY 1999 ANQENESTROYEGPLHYRQEQSPSPQOQ--LPPSSQAGMGQVPRTHRLITLADHIC 2056  
Db 181 ANQENESTROYEGPLHYRQEQSPSPQOQPLPPSSQSEGMQVPRTHRLITLADHIC 240  
QY 2057 QIITQDFARNQVSSQTPQPTSTFQNSPALSALVSTPVRTKTSNRYSPESQASVHHORPG 2116  
Db 241 QIITQDFARNQV---PSQPSSTSTFQNSPALSALVSTPVRTKTSNRYSPESQATVHLHPRPG 296  
QY 2117 SRVSPENLVDRSGSRGKSPERSHVSSEPYEPISPPQVPPVHVEKQDLSLLSQRGAEP 2176  
Db 297 PRVSPENLVDRSGSRGKSPERSHIPSEPYEPISPPQPAVHVEKQDLSMLLSQGVDP 356  
QY 2177 EORNDARSQGISYLPFFTKLENTSPWVSKKQIEFRKLNSGGGSDMAAAGPGEIF 2236  
Db 357 EORSDRSQGISYLPFFTKLESTSPWVSKKQIEFRKLNSGGGSDMAAAGPGEIF 416  
QY 2237 NLPVTTSGVSSSRGSHFADPASNGLGEDIIRKALMGSDFKVEDHGVMSQVHHORPG 2296  
Db 417 NLPVTTSGVSSSRGSHFADPASNGLGEDIIRKALMGSDFKVEDHGVMSHPVIMPGS 476  
QY 2297 ANTSVVTSGTTRREGDPSHSGGVCCKPKLISKNSRKSPIPGQYLGTERPSSVSV 2356  
Db 477 ASTSVVTSSEARRDEGEPSPHA-GVCKPKLINKNSRKSPIPGQYLGTERPSSVSV 535  
QY 2357 HSEGDYHROTQPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 2416  
Db 2417 NRIWEREPAPLSAQYETLSDSD 2440  
Db 596 NRIWEREPAPLSAQYETLSDSD 619

## RESULT 2

PCT-US95-16311-5  
; Sequence 5, Application PC/TUS9516311  
; GENERAL INFORMATION:  
; APPLICANT: Moore, David  
; APPLICANT: Seol, Wongi  
; APPLICANT: Choi, Hwang-Sik  
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street, Suite 3100  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16311  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,652  
; FILING DATE: 13-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/246001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 619 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; PCT-US95-16311-5  
  
Query Match 22.8%; Score 2885.5; DB 5; Length 619;  
Best Local Similarity 90.1%; Pred. No. 2.4e-171;  
Matches 562; Conservative 27; Mismatches 28; Indels 7; Gaps 3;  
  
QY 1819 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERKSV 1878  
Db 1 DALAALVDAASAPQMDVSKTKESKHEAARLEENLRSSAAVSEQQOLEQKNLEVERKSV 60  
QY 1879 QCLYTSSAFPSGKQPQSSVYVYSEAGDKGPPPKSRVYEEELTRGKTTITAANFIDVIT 1938  
Db 61 QCVCTSSALPSGKAQPHASVYVYSEAGDKGPPPKSRVYEEELTRGKTTITAANFIDVIT 120  
QY 1939 QRIASDKDAREGQSQSSSSLSHRYETPSDAIEVISPASSPAPPOEKLTQYQPEVVK 1998  
Db 121 QRIASDKDAREGQSQSSSSLSHRYETASDAIEVISPASSPAPPOEKQYQPDVVK 180  
QY 1999 ANQENESTROYEGPLHYRQEQSPSPQOQ--LPPSSQAGMGQVPRTHRLITLADHIC 2056  
Db 181 ANQENESTROYEGPLHYRQEQSPSPQOQPLPPSSQSEGMQVPRTHRLITLADHIC 240  
QY 2057 QIITQDFARNQVSSQTPQPTSTFQNSPALSALVSTPVRTKTSNRYSPESQASVHHORPG 2116  
Db 241 QIITQDFARNQV---PSQPSSTSTFQNSPALSALVSTPVRTKTSNRYSPESQATVHLHPRPG 296  
QY 2117 SRVSPENLVDRSGSRGKSPERSHVSSEPYEPISPPQVPPVHVEKQDLSLLSQRGAEP 2176  
Db 297 PRVSPENLVDRSGSRGKSPERSHIPSEPYEPISPPQPAVHVEKQDLSMLLSQGVDP 356  
QY 2177 EORNDARSQGISYLPFFTKLENTSPWVSKKQIEFRKLNSGGGSDMAAAGPGEIF 2236  
Db 357 EORSDRSQGISYLPFFTKLESTSPWVSKKQIEFRKLNSGGGSDMAAAGPGEIF 416  
QY 2237 NLPVTTSGVSSSRGSHFADPASNGLGEDIIRKALMGSDFKVEDHGVMSQVHHORPG 2296  
Db 417 NLPVTTSGVSSSRGSHFADPASNGLGEDIIRKALMGSDFKVEDHGVMSHPVIMPGS 476  
QY 2297 ANTSVVTSGTTRREGDPSHSGGVCCKPKLISKNSRKSPIPGQYLGTERPSSVSV 2356  
Db 477 ASTSVVTSSEARRDEGEPSPHA-GVCKPKLINKNSRKSPIPGQYLGTERPSSVSV 535  
QY 2357 HSEGDYHROTQPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 2416  
Db 536 HSEGDYHROTQPGAWEDRPSSTGSTQPPYNPLTIRMLSSPTPTIACAPSVAQAAPHQ 595  
QY 2417 NRIWEREPAPLSAQYETLSDSD 2440  
Db 596 NRIWEREPAPLSAQYETLSDSD 619  
  
RESULT 3  
US-07-741-940-2  
; Sequence 2, Application US/07741940  
; Patent No. 5352775  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY



APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/741,940  
FILING DATE: 19920109  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-741-940-2

Query Match 2.9%; Score 368; DB 1; Length 2843;  
Best Local Similarity 19.4%; Pred. No. 4.3e-14;  
Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

Qy 347 PEIRKQEQE-RQVRQCGAGLSATARSEH-----EISEIIDGLSQENNEK 395  
Db 981 PSIESYEDDESKEFCYQYPADLAHKIHSANHMDNDGELDTPTNYSLSKYSDQLNSGR 1040

Qy 396 QMRQLSVIPPMFDAQRRVKFINNGLMEDPMKVYKDRQPMNVTWTHE--KEIFKDKFI 453  
Db 1041 Q-----SPSONERWARPKHIIEDIEIKQEQQRQSRNOSTYYPVYTESTDDKHL 1087

Qy 454 QHPKNFGLIAYLERKSVDPCLVYYITKKNNYKALVRRNYGKRRGNQOIARPSQEK 513  
Db 1088 KFQPHFG-----QQECVSPY-----RSRGANGSETNRVGNHGINQVNSQLCOE- 1132

Qy 514 VEEKEDEKAERTEK--EEKKDEDEKEDSKENTKEDKIDGTAEETEEREQAPRG 571  
Db 1133 -DDYEDDKPTWYSEYSEEEHEERPTNYSIKYN-EERKHVDQPIDYSUKYATDIPSS 1190

Qy 572 R-----KTANSQGRKGRITRSTNEAAAS-----AAAAATEEPPLPPP 614  
Db 1191 KQSFSEKSSSGOSSKTEHSSSENTSTPSSNAKRNQNLHPSSAQSRSGQPQKAATCK 1250

Qy 615 PEPITSTPEVET-----SRWT-----EEMEVAKKGLVEHGRNWAIAKM--- 653  
Db 1251 VSSINQETIQTVCYEDTPICFSGSSLSLSAEDIGCNOTTOEADSANTLQIAEIKGK 1310

Qy 654 VGTKESE-----AQQKNFYNYKRRHNLNLLQOHK-----QKTSR 688  
Db 1311 IGRSAEDPVSEVPVSOHPRTKSRSLQGSLSSESARHKAVFPFGSAKSPSKSGAOTPK 1370

Qy 689 KPREE-----RDVSCQESVASTVSAQDEDEIAGNEBENPEDSEVAVKPS--DSPEN 740  
Db 1371 SPPEHYVQETPLMFRCVSV--SSLDSFESRSIASSVQSPCGSMVSGIISPDLDPSQ 1429

Qy 741 A--TSRGNTPEPAVELEPTTAPSTSPSLAVSTPKPADESVETOVNDISISAETAQMDV 798  
Db 1430 TMPFSPRSKTPPP--PPQATQTKREVPKNAKPTAEKRESGPKQAANAQV--RVQVLPA 1485

Qy 799 DQQHSAAE--GQVC-----DPPATKADSVDEYR-----VPEH--HASKVEGDN 840  
Db 1486 DTLHLFATESTPDGFCSSSISALSDEFFIQK-----DVELRIMPVQENDNGNETESQ 1541

Qy 841 TKE--RLDRLASEKVEPRDEDLVWAQINAORPEQSDNDSSATCSADEVD----- 890  
Db 1542 PKESNEQEKAEKTIIDSEKDLL-----DSD-----DDDIIECEIIS 1581

Qy 891 -----GEPERQRMF-----PMSDKPSLLNPTGSIILVSSPLKPNPLDLPLQHLRAAVI 937  
Db 1582 AMPTKSSRKGKPAQATASKLPPPVARKPSQL--PVYKLLPSQ-----NRL--QPQKHVSFT 1633

Qy 938 P----PWVSC--TPCNIPIGTPVSGYAL-----YQRHKAHME 969  
Db 1634 PGDDMPRYCYVEGTPINFSTATSLDLTIESPPNELAAGEVYRGGAQSGEFEKRDITPTE 1693

Qy 970 SAL-----LEEQRORQEQIDLECRSSSTSPCGTSKSPNREVL----- 1007  
Db 1694 GRSTDEAOGGKTSSVTIPELDDNKAEGDILAEICNSAMPKGSKHPKRVKIMDOVOQA 1753

Qy 1008 ---QAPHQLITNLPPEGVRLPTRPRPPPLI-----P 1038  
Db 1754 SASSSAPNK--NQLDGKKKPTSPVKIPQNTYRVRKNADSKNNLNAERVSNDKD 1810

Qy 1039 SSKTTVASEKPSF-----IMGGSIS-----QGTPTGYLTLSHNOASTQETP 1079  
Db 1811 SKQNLKNNKDFNDKLPNNEDVRGSEAFSPHHTPIEGTP--YCFSRND----- 1860

Qy 1080 KPSVGSISLGLPROQESAKSATLVYKQEEFSPRSQSQPEGLLVRAQHEGVVVRGTAGAI 1139  
Db 1861 --SLSSLDLDDDDVLDLSREKAEELKAKENKESEAKVTSHT--LTSNOOSA---NKTQAI 1913

Qy 1140 QEGSITRGTPTSKISVESIPSLRGSITOGTPALPOTGPTPTALVK-----G 1185  
Db 1914 AKQPINRGQPKPILQKQS-----TFQSSKIDIPRGAATDEKLQNAFIENTPVCFSHNS 1967

Qy 1186 SISRM-----PIEDSSPEKGRERAAKSGHVIYBKGSGHI----- 1219  
Db 1968 SLSSLDIDQNNNNKENENEPKETETPPDSQGEPSKPKQASGYAPKSPHVEDTPVCFSRNSSL 2027

Qy 1220 --LSYDN-----IKNAREGTRSPRTAHEISLRKSYSEVGNIKQGMSPRESVPSAPL 1269  
Db 2028 SLSLSDSEDDLLOECISSAMPKKKPSRLKGDNEKHSPRNMGGLGEDLTLDLKDQRPD 2087

Qy 1270 EGLICRALPROGSPHDLKERTVLSIMQGTTPRATTEFEDGLKYPKQIKKRESPIRAFE 1329  
Db 2088 SE--HGLSPDSENFDMKAIQIEGANSIVSSLHQAAAAA-----LSRQASSDSLSILK 2139

Qy 1330 GAITKGKPYDGIITTIKEMGRSITHEIPRODILTQESRKTPVYVQSTRPIEGSISOGTPIK 1389  
Db 2140 SGISLGSFP-----HLTPDQEEKPFTSNKGPRILK-----PGEKSTLETKK 2180

Qy 1390 PDNNS--GOSAIKHNKVSILITGP-----SKLSRGM--PPLIIVPENIKVVRGKYEDYKAGET 1443  
Db 2181 IESESGIKGKGVKVSILITGKVRNSNISEISGMKQPLQ---ANMPSISRG----- 2227

Qy 1444 VRSRHTSVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSYONTMWSRGSPMMNRTS 1503  
Db 2228 -----RTMIH-----IPGV-----RNSSSSTSPVSKRGP 2251

Qy 1504 DVTIPPNNKSTNHERKSTLTPT--QRESIPAK--SPVPGVDPVYVSHSPFDPHHRGSTAGEVYW 1561  
Db 2252 PLKTPASKSPSEGGTATTSPGAKPSVKSELSUPVARQTSQIGGSSKAPSRSGSR----- 2305



```
QY 1562 SHLPQLDPAFFHRAIDPAAAYLFQRLSPTPGYPSQYQIYAMENRQTIILNDIITSQ 1621
Db 2306 -----DSTPSRPAQQL-----SR 2319
QY 1622 QMOVNLRPDVA---RGLSPREOPLGLPYPAIRGIIDLINMPTTILVPHPGGTSPMDRI 1678
Db 2320 PIOSGRNISIPGRNGISPPNK-----LSQLPRT-SSPSTASTKSGSGKM 2364
QY 1679 TVIPQTQTFFPPYNSASMSGPHPTLHAAAASABRERERE-----KERER 1726
Db 2365 SY-----TSGRQMSQNLIT--KOTGLSKNASSIPRESASKGLNQMNNGANKVKEL 2416
QY 1727 ERIAAASDLRLPGEQGRPGSHGYVRSPSPSVRTQETMLQORPSVFOGTNGTSTVTP 1786
Db 2417 SRMSSTKSGS-----GSED-----RSERPVLRQSTFIKEAPS-----2449
QY 1787 LDPTAQLRIMPLPAGGPISOGLPASRYNTAADALUALDAASAPQMDVSKTKESKHEA 1846
Db 2450 --PTLR-RKLEESASFELS---PSSRPASPTRSQAQTPVLSPLDMSLS-THSSVQAG 2502
QY 1847 A--RLEENLRSRSAVSEQOOLEOKTLEKRSVQCLYTSSAFPSGKPOPHSSVYVSEAG 1904
Db 2503 GWRKLPPNL---SPTIENDGRPAKHDIAR-----SHSESPRLPINSRGTWKREHS 2552
QY 1905 KDKGPPPKSRYEELTRGKTTITAAAN-----FIDVIITRIQIASDKDARERGQSSD 1956
Db 2553 KHSSSLP--RVSTWRTGSSSILSASSESEKAKSEKHKVNSISGTKQSKENQVSAKG 2610
QY 1957 SSSLSLSHYE--TPSDATEVISPASSAPPOEKLQTOPEVVKANQAENDPTROVEGLPH 2015
Db 2611 TWRIKIKENFEFTNSTQTSVSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPIN 2667
QY 2016 HYRPOQESPPQOQLPP-----SSQAEQMGQVPRTHRLITLADHI 2055
Db 2668 NRP---SGRSTGNTPPVIDSVSEKANPNIKDNDQNAKQNGVSGVPM--RTVGLENRL 2722
QY 2056 CQIITQDFARNQVSSQTPOQPT--STFQNSPSALVST---PVRTKTSNRYSPESQAQSV 2110
Db 2723 TSFIQVD-----APDQKGTEIKPGONNPVPVSETNESPIVERT-----PFSSSSSS 2768
QY 2111 HHQRP---GSRVSPENLVDSRGRPKSPERSHVSSEPEYPTSPPOVPVH--EKODS 2164
Db 2769 KHSSPGTVAARVTPFNY-----NPSPRKSSADTSARPSQIP--TPVNNNTKKRDS 2818

RESULT 4
US-08-289-548A-2
; Sequence 2, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allogretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-289-548A-2
```

```
Query Match 2.9%; Score 368; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 4,3e-14;
Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;
```

```
QY 347 PEIRKORQOE-RQRYQRCAGLSATARSEH-----EISEIIDGLSEQENNEK 395
Db 981 PSIESYDEDDSKCSYQYPADLAHKHTSANHMDNDGELDTPIYSLKYSDEQLNSGR 1040
QY 396 QMRQLSVIPPMFADQRRVFINNGLMDEPMKYVKDRQFMVNTDHE--KEIFKDFI 453
Db 1041 Q-----SPQNERWARKHIIIEDEIKQSEQRQNRQSTTYPVTTESTDDKHL 1087
QY 454 QHPKNFGLIASYLERKSVDCVLYLYLTKNENYKALVRRNYKRRGRNQIARPSQEK 513
Db 1088 KFQPHFG-----QOECVSPY-----RSRGANGSETNRVGNHGINQVNSQLCOE- 1132
QY 514 VEEKEEDKAEKTEK--EEKKDEEKEDEKEDSKENTKEKIDGTAETETEREQATPRG 571
Db 1133 -DDYEDDKDPTWYSERYSEEEQHEERPTNYSIKYN-EKRRHVDPQIDYSLKYATDIPSS 1190
QY 572 R-----KTANSQGRKRGRITRGMTEAAAAA-----AAAAAATEEPPLPPP 614
Db 1191 QKQSFESKSSGSSGSKTEHMSSESSENTSTPSSNAKRNQLHPSSAQSRSGQPKAATCK 1250
QY 615 PEPITSTEPVET-----SRWT-----EEMEVAKKGLVEHGRNAAIAKM---653
Db 1251 VSSINQETIQTVCYVEDTPICFSGSSLSLSSAEDEICGNTTOEADSANTLOIAEIKGK 1310
QY 654 VGTKSE-----AQCKNFYNYKRRHNLNLLQHK-----QKTSR 688
Db 1311 IGRSAEDPVSEVPVAVSOHPRTKSRLOGSSLSSESARHKAVEFPSSGAKSPKSGAQTPK 1370
QY 689 KPREE-----RDVSOQESVASTVSAQDEDEIDIASNEENEPDESEVAVKPSK--DSPEN 740
Db 1371 SPPEHYVQETPLMESRCTSV--SLDSFESRSIASVQSEPCSGMVSGIISFDLPDSGQ 1429
QY 741 A--TSRGNTPEPAVELEPTTETAPSTPSLAVPTKPADESVEVQVNDISISAETAQMDV 798
Db 1430 TMPPSRSKTPPP---PPQATQTKREVKNKAPTAEKRESGPKAAVNAVQ--RVQVLPDA 1485
QY 799 DQOESHAEE---GSVC-----DPPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840
Db 1486 DTLHLFATESTPDGFCSSSSLSLSLDEPFTQK---DVELRIMPVQENDNGNETESEQ 1541
QY 841 TKE--RDLDRASEKVEPRDELDVVAQQINAOQRPESQDNDSSATCSADEVD-----890
Db 1542 PKESNQEKEAEKTIIDSEKDL-----DDSD-----DDDEILEECIIS 1581
QY 891 -----GEPERQRMF-----PMDSKPSLLNPTGSIIVSSLPKPNLDLPQHQHRAVI 937
Db 1582 AMPTKSSRKGGKKPAQTASKLPPPVARKPSQL-PVYKLLPSQ---NRL---QPOKHVSFT 1633
```



Qy	938	P----	PWVS-	---TPONIPIGTPVSGYAL-----	-----YQRHIKAMHE	969						
Db	1634	PGDDMPRVYCEGTPTINFSTATSLDLTIESPPNELAAGEVGRGAQS	GEPEKRD	TIPT	1693							
Qy	970	SAL-----	LEQRORQ	QOIDL	ECRSSTSPCGTSKSPNREVL-----	1007						
Db	1694	GRSTDEA	QGGKTS	SVTIP	ELDNDK	AEEDGILAE	CINSAMPK	GSHK	FRVK	KIMD	QVOQA	1753
Qy	1008	--OPAPHOLITNL	PEGVRLPTTRTRPPPLI-----	-----P	1038							
Db	1754	SASSAPNK--	NOLDGKKK	KTSPVKPI	PQNTERTY	TRVKN	ADSKNNL	NAER	VDSN	DK	1810	
Qy	1039	SKTTT	VASEK	PSF-----	IMG	STS-----	QCTPGTY	LTSH	NOASY	TOETP	1079	
Db	1811	SKKQNL	KNNSK	DFNOK	LPN	NER	VRG	SFA	DSPPH	YHTPI	E	1860
Qy	1080	KPSVG	SISL	GLPRO	QESAK	SATL	PIYIK	QEE	SP	SONSQ	PEGL	1139
Db	1861	--SLSD	DFDD	DDV	LSR	KAEL	RKAKEN	KESEAK	VTSHTE--	ITS	NOOSA---	1913
Qy	1140	QEGSI	TRG	PTSKIS	VES	PS	LRG	STIG	TOT	PAL	POT	1185
Db	1914	AKQ	INR	GPK	ILQ	KOS-----	TFPOSS	KDIP	DRGA	ATDEK	LQ	1967
Qy	1186	SISM-----	PIED	SP	KEG	REAA	SKGH	VI	Y	ECK	SGHI-----	1219
Db	1968	SLSL	SDID	QENN	KNEN	EPIKET	EP	DSQ	GP	SKPQ	ASGYAP	2027
Qy	1220	--LSYDN-----	IKNA	REG	TRSP	TAHE	IS	LS	KRS	VE	SGNI	1269
Db	2028	SSLIS	DE	DD	LQ	EC	LSS	AMP	KKK	PSRLK	GDNE	2087
Qy	1270	EGLIC	ALPR	GSP	HD	KL	ERT	VL	SG	IM	QGT	1329
Db	2088	SE---	HGL	SP	DE	N	ED	WKAI	Q	E	ANS	2139
Qy	1330	GAT	TK	KPY	D	G	IT	T	IK	EM	G	1389
Db	2140	SGIS	LG	SPF-----	HLTP	DQ	E	K	P	FTS	NK	2180
Qy	1390	FDNNS	-Q	SA	I	KN	Y	K	SLIT	GP-----	SKLS	1443
Db	2181	IES	E	KG	I	G	K	G	K	V	Y	2227
Qy	1444	VR	S	R	H	T	S	V	S	S	G	1503
Db	2228	-----	RTMI	H-----	IPGV-----	-----	RN	SS	S	T	S	2251
Qy	1504	D	V	T	P	N	K	T	N	H	E	1561
Db	2252	PLK	T	P	A	K	S	P	S	E</		

RESULT 5  
 US-08-452-654-2  
 ; Sequence 2, Application US/08452654  
 ; Patent No. 5691454  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALBERTSEN, HANS  
 ; APPLICANT: ANAND, RAKESH  
 ; APPLICANT: CARLSON, MARY  
 ; APPLICANT: GRODEN, JOANNA  
 ; APPLICANT: HEDGE, PHILIP J.  
 ; APPLICANT: JOSLYN, GEOFF  
 ; APPLICANT: KINZLER, KENNETH  
 ; APPLICANT: MARKHAM, ALEXANDER F.  
 ; APPLICANT: NAKAMURA, YOSUKE  
 ; APPLICANT: THLIVERIS, ANDREW  
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 ; NUMBER OF SEQUENCES: 94  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Banner, Birch, McKie & Beckett  
 ; STREET: 1001 G Street, NW  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20001-4598  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/452,654  
 ; FILING DATE: 25-MAY-1995  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/741,940  
 ; FILING DATE: 08-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kagan, Sarah A.  
 ; REGISTRATION NUMBER: 32,141  
 ; REFERENCE/DOCKET NUMBER: 1107.035574  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-508-9100  
 ; TELEFAX: 202-508-9299  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2843 amino acids  
 ; TYPE: amino acid



TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-452-654-2

Query Match 2.98; Score 368; DB 1; Length 2843;

Best Local Similarity 19.4%; Pred. No. 4.3e-14;

Matches 418; Conservative 281; Mismatches 797; Indels 664; Gaps 98;

QY 347 PEIRKQROQE-RQVRCORAGLSATIRSEH-----EISEIIDGLSQENNEK 395  
DB 981 PSIESYDEDESKFCYGOYPADLAHKIHSANHMDDNDGELDTPTINSLKYSDQELNSGR 1040  
QY 396 QMRQLSVIPPMFAEORRVFIMNGLMEDPMKYKROFMNYWTIDHE--KEIFKDFI 453  
DB 1041 Q-----SPSONERWAPKHIIEDEIKOSEQRQSRNSTTYPVVTSTDDKHL 1087  
QY 454 QHPKNFGLIASYLERKSPDCVLYLTTKKNENYKALVRRNYKRRGRNQOIARPSOEK 513  
DB 1088 KFQPHFG-----QQECVSPY-----RSKGANGSETNRVGSNHGINQNVSQLQE- 1132  
QY 514 VEEKEEDKAEKTEK--EEKKDEEKDEKEDSKENTKEKDKIDGTABETEERQATPRG 571  
DB 1133 -DDYEDDKPTNYSERYSEEEQEEERPTNYSIKYN-EEKRHVDPQIDYSLKYATDIPSS 1190  
QY 572 R-----KTANSORRRKGRITRSMNTNEAAS-----AAAAATEEPPLPPP 614  
DB 1191 QKQSFPSKSSGSSKSTEHMSSSENSTSPSSNAKRONQLHPSSAQSRQSQPOKATCK 1250  
QY 615 PEPISTEPVET-----SRWT-----EEMEVAKKGLVEHGRNWAIAK-- 653  
DB 1251 VSSINQETIQYCVEDTPICFSRGSSLSLSSAEDEICGNTQOEADSAANTLQIAEIKG 1310  
QY 654 VGTKE-----AQCKNFYKRRHLDNLQOQK-----QKTSR 688  
DB 1311 IGRTSAEDPVSEVAVSHQPTKSRQGLSSSESARHKAVERFPFGAKSPSKSGAOTPK 1370  
QY 689 KPREE-----RDVQSCEVASTVSAQDEIDAEANEENEDSEVEAVKPSE--DSPEN 740  
DB 1371 SPPEHYVQETPLMFSRCTSV-SLSDSFERSIASVQSEPCSGMVSGIISFSDLPDSPG 1429  
QY 741 A--TSRGNTPEVALEPTETAPSTPSLAVPTKPADESVETQVNDISAEATAEQMDV 798  
DB 1430 TMPSPRSKTPPP--PPQTAOTKREVKNKAPTAEKRESGPKQAAVNAVQ-RVQVLPA 1485  
QY 799 DQEHSAEE--GSVC-----DPPATKADSDVDEVR-----VPEN-HASKVEGDN 840  
DB 1486 DTLHFATSTPDGFCSSLSLSLDEPFIQK-----DVELRIMPPVQENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDEDLVVAQINAQRPESQDNDSSATCSADEVD----- 890  
DB 1542 PKESNEQKEAETIDSEKDLL-----DDSD-----DDDIIEILEECIIS 1581  
QY 891 -----GPERQRMF-----PMDSKPSLLNPTGSLVSLPKPNLDLPQLOHRAVI 937  
DB 1582 AMPTKSSRKPKAQTASKLPPPVARKPSQL-PVYKLLPSQ-----NRL-----QPQKHVSFT 1633  
QY 938 P----PMVSC---TPCNIPICTPVSGAL-----YQRHKAMHE 969  
DB 1634 PGDDMPRYVCEGTPINFSTATISLSDLTIESPPNELAAGEVGRGAOSGEFEKRDITPTE 1693  
QY 970 SAL-----LEORQROEQIDLECRSSTSPCGTSKSPNREWEVL----- 1007  
DB 1694 GRSTDEAOGGKTSVVTIPELDNDKAEEDGILLAEICINSAMPKPKSHKPRVYKKINDVQQA 1753  
QY 1008 ---OPAPHOLLTNLPEGVRLPTTRPPPLI-----P 1038  
DB 1754 SASSAPNK---NOLDGKKKPTSPVKPIQNTYEYTRVRKNADSKNNLNAERYFSDNKD 1810  
QY 1039 SSKTTVAEKPSF-----IMGGTS-----CGTPTYLTSHNOASYTOETP 1079  
DB 1811 SKQNLRKNSKDFNDKLPNNEDVRGSAFDSPHYPIEGTP--YCFSRND----- 1860

QY 1080 KPSVGSISLGLPROQESAKSATLPYIKOEFSRSONSQPEGLLVRAQHEGVVRGTAGAI 1139  
DB 1861 --SLSLDFDDDDVDLSREKAELKAKENKESEAKVTSHTL--LTSNQOSA--NKTQAI 1913  
QY 1140 QEGSITRGTPTSKISVESIPSLRGSITOGTPALPOTGIPTALVK-----G 1185  
DB 1914 AKQPINRGQPKPILQKQS-----TFQSSKDIIDPRGAATDEKLFQNFATIENTPVCFSHNS 1967  
QY 1186 SISRM-----PIEDSSPEKREEAASKGHVIECKSHI----- 1219  
DB 1968 SLSLSDIDQENNNKENPEPIKETEPDQSGEPKQASGYAPKSPHVEDTTPVCFSRNSSL 2027  
QY 1220 --LSYDN-----IKNAREGTRSPRTAHSILSKRSVESVEGNIKQGMRESVPSAPL 1269  
DB 2028 SLSIDSEDDLLQECISSAMPKKKPSRLKGDNEKHSRNMGGILGEDLTLDLDIQRPD 2087  
QY 1270 EGLICRALPRGSPHSDUKERTVLSGSIQGTTPRAATTESFEDGLKYPQIKRESPIRAFE 1329  
DB 2088 SE--HGLSPDSENFDMKAIQEGANSIVSSLHQAAAAAC-----LSRQASDSDSILSK 2139  
QY 1330 GAIYKGRPYDGIITIKEMGRSIHEIPRODILTQESRKTPEVQSTRPIEGSISQGTPIK 1389  
DB 2140 SGIISLGSF-----HLTPQOEKPFSTSNKGPRILK-----PGEKSTLETKK 2180  
QY 1390 FDNNS-GQSAIKHNKSLITGP---SKLSRGM-PPEIVPENIKVVERGYEDVKAGET 1443  
DB 2181 IESEKGLKGGKVKYKSLITGKVRNSSEISQMKQPLQ---ANPMSISRG----- 2227  
QY 1444 VRSRHTSVVSSGSPVLSTLHEAPKAQLSPGIYDDTSARRTPVSYQNTMTRSGSPMMNRTS 1503  
DB 2228 -----RTMIH-----IPGV-----RNSSSSTSPSVKKG 2251  
QY 1504 DVTTPNKNSTNHERKSLTPT-QRESIPAK-SPVPGVDVPSHSPFDPHHSRGTAGEVYW 1561  
DB 2252 PLKTPASKSPSEGOTATSPRGAKPSVKSELSVARQTSQIGGSKKAPSRGSR----- 2305  
QY 1562 SHLPTQLDPAMPFFHRALDPAAAYLFORQLSPTPGYSQYQIYAMENTROTILNDYITSQ 1621  
DB 2306 -----DSTPSRPAQQL-----SR 2319  
QY 1622 OMQVNLRPDVA---RGLSPREQPLGLPYPATRGIIDLTMPPPTILPHPGGTSTPPMDRI 1678  
DB 2320 PIQSPGRNISIPGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGSKM 2364  
QY 1679 TYIPGTQITTEPPRYNSASMSGPHPTHLAAASAARERERE-----KERER 1726  
DB 2365 SY-----TSPGRMSQONLT--KOTGLSKNASSIPRSESASAKGLNQNNNGANGANKVEL 2416  
QY 1727 ERIAAASDLYLRPGSEQPGRPGSHGYVRSPSPSVRTQETMLOORPSYVFOGTNGTSVITP 1786  
DB 2417 SRMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS----- 2449  
QY 1787 LDPTAQLRIMPLPAGGSPISQGLPASRYNTAADALAAALDAASAPQMDVSKTESKEHA 1846  
DB 2450 --PFLR-KLEESASFESLS---PSSRPASPTRSQATPVLSPLPDMSLS-THSSVQAG 2502  
QY 1847 A--RLEENLRSRSAVSEQQOQLEKQTLVEKRSVQCLTSSAFTSGPKPOPHSSVYSAG 1904  
DB 2503 GWRKLPLNL---SPTIYNDGRPAKRHDIAI-----SHSESFRSLPINKSGTWKREHS 2552  
QY 1905 KDGPPPKRSYEEELRTGKTTITAAN-----FIDVIITRQIASDKDAREGSSQSD 1956  
DB 2553 KHSSSLP--RVSTWRRRTGSSSSILSSASSESESEKAKSEKDEKHNVSIGTKQSKQSNVSAK 2610  
QY 1957 SSSSLSSHRYE-TPSDAIEVISPASSAPPQEKLOTQPEVVKANQAENDTTRQYEGPLH 2015  
DB 2611 TWRKIKENEFSPTNSTOTVSSGATNGAESKTLIQMAPAVSK---TEDVWVRIEDPCIN 2667  
QY 2016 HYRQOQESPPQOQLPP-----SSQAEQMGQVPRTHRLITLADHI 2055  
DB 2668 NPR---SGRSTGTWTPVIDSVSEKANPNIKDKDNQAKQNVGNSVPM--RTVGLENRL 2722  
QY 2056 CQIITQDFARNQVNSQTPQOPPT--STFQNSPSALVST---PVRTKTSNRYSPESQAOSV 2110







Db 2181 IESESGIKGGKVKYKSLITGKVRNSSEISGMKQPLQ---ANMPSISRG----- 2227  
QY 1444 VRSRHTSVSSGSPVLRSTLHEAPKAQLSPGIYDDTSARRTPVQVQNTMSRSGSPMMRTS 1503  
Db 2228 -----RIMIH-----IPGV-----RNSSTSPVSKKGP 2251  
QY 1504 DVTIPPNKSTNHERKSTLPT-QRESIPAK-SPVPGVDVSVHSPFDPHHRGSGAGEVYW 1561  
Db 2252 PLKTPASKSPSEGATSPRGAPSKVSELSPVARTSQIGGSKAPSRSGR----- 2305  
QY 1562 SHLPTQLDPAMPFRHALDPAAAVLFORQLSPTPGYSQVQLYAMENRTOTILNDYTSQ 1621  
Db 2306 -----DSTPSRPAQOPL-----SR 2319  
QY 1622 QMOYNLRPDVA---RGLSPREPGLGLYPATRGIIIDITNMPPTLIVPHPGCTSTPPMDRI 1678  
Db 2320 PIQSPGRNSISPGNGISPPNK-----LSQPRT-SSPSTASTSSGSGKM 2364  
QY 1679 TYIPGTQITPPRPYNASMSPGHPTHLAAAASAEERERERE-----KERER 1726  
Db 2365 SY-----TSPGRQMSQNLTI-KQTGLSKNASSIPRSESASKGLNQMNNGANGANKVEL 2416  
QY 1727 ERIAAASDLVLRGSGOPRPGSHGVSPSPSVRTQETMLQORPSVFOGTNGTSTVTP 1786  
Db 2417 SRMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS----- 2449  
QY 1787 LDPTAQLRIMPLPAGGSPISQGLPASRYNTAADALAAALVDAASAPOMDYKTESKHEA 1846  
Db 2450 --PLR-RKLEESASFELS---PSSRPASPTRSQATPVLSPPLDMSLS-THISSVQAG 2502  
QY 1847 A--KLEENLRSAVAESEOQLEKTELEKRSVQCILYTSAPPSGKPPHSSVYVSEAG 1904  
Db 2503 GWRKLPPNL---SPTIEYNDGRPAKRDHDIAR-----SHSESPSLPINKSGTWKREHS 2552  
QY 1905 KDKGPPPKSYEEELRTRGTTITAAAN-----FIDVITITRIASDKDARERGSOSSD 1956  
Db 2553 KHSSSLP--RVSTWRRGTSSSILUSSESEKAKSEKHNVSISGTSKOSKENQVSAKG 2610  
QY 1957 SSSSLSHRYE-TPSDAIEVISPASSPAPPOEKLTQYQPEVVKANQANENDPTROYEGLPH 2015  
Db 2611 TWRIKIKENESPTNSTSOTVSSGATNAESKTLIYQMAPAVSK---TEDVWVRIEDCPIN 2667  
QY 2016 HYRQOQSPSPQOOLPP-----SSAEGMGVQVPRHRLITLADHI 2055  
Db 2668 NPR---SGRSPTGTPPTVIDSVSEKANPNIKDSKDNQAKQVNGSVPM--RTVGLENRL 2722  
QY 2056 CQIITQDFARNOVSSOTPOQPPT--STFQNSPALVST--PVTKTSNRYSPESQAQSV 2110  
Db 2723 TSFIQVD-----APDQGTGIKQGNPVPVSETNESPIVERT-----PFSSSSS 2768  
QY 2111 HHQRP---GSRVSPENLVKSRGSRPKGSPERSHVSSEPEYPISSPQVPVH--ERQDS 2164  
Db 2769 KHSSPSGTVAARVTPFNY-----NPSRKSSADSTARSQIP-TPVNNNTKKRDS 2818

## RESULT 7

US-07-741-940-7

Sequence 7, Application US/07741940

Patent No. 535275

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/741,940  
; FILING DATE: 19920109  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.035574  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2842 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: APC  
; US-07-741-940-7

Query Match 2.9%; Score 365; DB 1; Length 2842;

Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PETRKOREQE-RFQVQRGAGLSATIRSEH-----EISEIIDGLSEQENNEK 395  
Db 980 PSIESYSEDESEFCSYQYPADLAHKHISANHMDDNGELDTPIYSLKYSDEQLNSGR 1039  
QY 396 QMRQLSVIPMMFDAEORRYKFINMGLMEDPMKYKDRQFMVWTDHE--KEIFKDKFI 453  
Db 1040 Q-----SPQNERWARPKHIIIEDEIKQSEQRSRNQSTIYVYTESTDDKHL 1086  
QY 454 QHPKNFGLIASYLERKSVDPCLVLYITKKENYKALVRRNYKRRGRNQOIARPSQEEK 513  
Db 1087 KFPQHFQ-----QCECVSPY---RSRGANGSETNRVGSNHGICNVQSQCQE- 1131  
QY 514 VEEKEEDKAEKTEKK--EEBKDEBEKDEKEDSKENTKEKDKIDGTAEETEEREQATPRG 571  
Db 1132 -DDYEDDKPTNYSERYSEEEQHEEERPTNYSIKYN-EKRRHVDQPIDYSLKYATDIPSS 1189  
QY 572 R-----KTANSQGRKGRITRSMTNEAAS-----AAAAATEEPPPLPPP 614  
Db 1190 OKOSFESKSSGSGSKTEHMSSESSTSTPSSNAKQNLHPSSAQSRSGQPKAATCK 1249  
QY 615 PEPITPEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAIAK 653  
Db 1250 VSSINQETIQTVCVEDTPICFSCSSLSLSAEDIGCNOTTQEADSANTLQIAEIKK 1309  
QY 654 VGTKSE-----AOCKNFYNTKRRNLNLOOHK-----OQTSR 688  
Db 1310 IGRSAEDPVSEVPVAVSQHPRTKSSRLQGLSSLSSEARHKAVERFSSGAKSPSKGATPK 1369  
QY 689 KPREE-----RDVSCQESVASTVSAQEDIEDIASNEENPEDEVEAVKPS--DSPEN 740  
Db 1370 SPPEHYVQETPLMFSRCTSV-SSLDSFESRSIASSVQSEPCSGMVSGIISFSLDPSFGQ 1428



QY 741 A--TSGNTEPAVELEPTTAPSTSPSLAVPSTKPAEDSVETQVNDISIAETAQMDV 798  
DB 1429 TAPPKSRKTPPP---PPQATQTKREVPAKNKAPTAKREKSPKQAAVAAVQ-RVQVLPDA 1484  
QY 799 DQOEHSAEE---GSCV-----DPPATKADSVDEVR-----VPEH-HASKVEGDN 840  
DB 1485 DTLHFATFESTPDGFCSSSSLSALSLEDFIOTK-----DVELRIMPVQENDNGNETESEQ 1540  
QY 841 TKE--RDLDRASEKVEPRDELDVVAQINAQRPQSDNDSSATCSADEVD-----890  
DB 1541 PKESNENGEAEKIDSEKOLL-----DSD-----DDIEIEECILIS 1580  
QY 891 ---GEPERQRMPPMDKPSLLNPTGSSILVSPKPNPLD-----LP-----QLQHRAAVIP 938  
DB 1581 AMPTKSSRKAKKPAQATKSLPPVAR-----KPSQLPVYKLLPSQNLQFQKHVSFT 1633  
QY 939 ---PMVSC---TPCNIPGTPVSYAL-----YQRIKAMHES 970  
DB 1634 GDMPRVYCVGTPINFSTATSLSDLTIESPPNELAAGEVGRGAQSGEFKRDPTIPEG 1693  
QY 971 AL-----LEQORQEOIDLECRSSTSPCGTSKSPNREWEVL-----1007  
DB 1694 RSTDAQGGKTSSTVITPELDONKAEEDILAEICINSAMPKSKHKPFVKKIMDOVQOAS 1753  
QY 1008 --QPAPHOLITNLPVGLPRTTRPPPLI-----PS 1039  
DB 1754 ASSAPNK---NOLDGKKKPTSPVKPIQNTVEYRVRKNADSKNNLNAERVFSDNKDS 1810  
QY 1040 SKTTVASEKPSF-----IMGSSIS-----OGTPGYLTSNHQASTYQETPK 1080  
DB 1811 KKQNLKNNKDFDKLPNNEDRVGSAFADSPHHYPTIEGTP--YCFSRND-----1859  
QY 1081 PSVGSISLGLPROQESAKSATLPIKQEFSPRSONSQPEGLLVRAQHEGVVGTAGAIQ 1140  
DB 1860 -SLSLDDDDVLSRKAELRAKENKESEAKVTSITE--LTSNQOSA---NKTQAI 1913  
QY 1141 EGSITRGTPTSKISVESIPSLRSGITQTPALPQTIPTALVK-----GS 1186  
DB 1914 KQINRGQPKPILOKS-----TFQSSKDIPOGAAATDEKLQNFALENTPVCFSHNS 1967  
QY 1187 ISRM-----PIEDSSPEKGREAAKSHVIEKSGHI-----1219  
DB 1968 LSSLSDDIQENNKENEPKETEPDQSGEPSKPAQYAPKSFHVEDTPVCFSRNSSL 2027  
QY 1220 -LSVDN-----IKNARECTSPRTAHEITSLKRSYSEVEGNKOGMSRSPVSAPLE 1270  
DB 2028 SLSIDSEDLLOECISSAMPKKKPSRLKGDNEKHSRPMGIIIGEDLTLQDKIQRPDS 2087  
QY 1271 GLICRALPGSPHSDLKERTVLSGIMQTPRATESPEDGLKYPKQIKRESPPIRAPEG 1330  
DB 2088 E---HGLSPDSENFWDKAIQEGANSIVSSLHQAAC-----LSRQASSDSLSLSKS 2139  
QY 1331 AITGKPKVDGTTIKEMGRSTHEPRQDILTQESKRPVVOVQTRPIIEGSIQGTPIKF 1390  
DB 2140 GISLGSPP-----HLTPDQEKPFNTKNGPRILK-----PGEKSTLETKKI 2180  
QY 1391 DNNS-GOSAIKHNVKSLITGP-----SKLSRGM-PLEITVPENIKVVERKYEDVKAGETV 1444  
DB 2181 ESEKIGKGGKVKYKSLITGVKRNSELSGQMKOPLQ---ANMFSIRG-----2226  
QY 1445 RSRTSVVSSGSPVLSRSLTHLPAKQASPGIYDDTSARPTVSYQNTMSRGSPPMNRSTD 1504  
DB 2227 -----RTMHI-----IPGV-----RNSSTSPVSKGKGP 2251  
QY 1505 VTIPNKNSTNHERKSTLPT--QRESIPAK-SPVPGVDVPSHSPDPHHRSGTAGEVVWS 1562  
DB 2252 LKTPASKSPSEGOATTSPRGAKPSKSELSPVARQTSIQIGSSKAPSRSGSR-----2304  
QY 1563 HLPTQLDPAMFPHRALDPAAAYLFQRLSPTPGYPSQYQLYAMENTROTILNDYITISQ 1622  
DB 2305 -----DSTPSRPAQOPL-----SRP 2319  
QY 1623 MQVNLRPDVA---RGLSPREQLGLPYPATRGIIDLTMNPPTILVPHPGGTSTPPMDRIT 1679

DB 2320 IQSPGRNISPCRNIGISPKNK-----LSQLPRT-SSPSTASTKSSGSGKMS 2364  
QY 1680 YIPGTOITFPFPPPNASMSGPHTHLAAAASABERERERE-----KERERE 1727  
DB 2365 Y-----TSPGRQMSQOQLT--KOTGLSKNASSIPRSESASKGLNMNGNGANKKVELS 2416  
QY 1728 RIAAASSDLYLRPQSEQPRGSHGYVRSPSVRTQMLQORPSVFOGTNGTNGTSVITPL 1787  
DB 2417 RMSSTKSS-----GSESD-----RSEPLVLRQSTFFKEAPS-----2448  
QY 1788 DPTAQLRIMPLPAGGPSISOGLPASRYNTAADAALAYDAASAAPQMDVSKTKESKHEAA 1847  
DB 2449 -PTLR-RKLEESASFESLS---PSSRPASPTRSQAQTQVLSPLPDMSLT-THSSVQAGG 2502  
QY 1848 --RLEENLRSAVSEOOQLEQKTLVEKRSVQCLYTSSAFFPSGKPOPHSVVYSEAGK 1905  
DB 2503 WRKLPPNL---SPTIYNDGRPAKRHDIAI-----SHSESPSLPINRSGTWKREHSK 2552  
QY 1906 DKGPPPKSRYEELRTRGKTTITAAI-----FIDVITRQIASDKDARERSQSSDS 1957  
DB 2553 HSSSLP--RVSTWRTGSSSILSASSSESEKAKSEDEKHVNSISGTQSKENQVSAKGT 2610  
QY 1958 SSSLSHRYE-TPSDAIEVISPASSPAPPQBEKLOTYQEVVKANQAENDPTRQYEGPLHH 2016  
DB 2611 WRKIKENEFSTNSTQTSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2667  
QY 2017 YRQOESPSPOQOLPP-----SSQAEQMGQVPRTHRLITLIADHIC 2056  
DB 2668 PR---SGRSPGTGNTPPVIDSVSEKANPNIKDKONQARQNGVSGVPM--RTVGLENL 2722  
QY 2057 QIITQDFARNQVSSQTP-QOQPSTSTFONSPSALVS-TPVRTKTSNRYSPESQAOQSVHHQR 2114  
DB 2723 SFIQVDADQKTEKPKQNNPVPVSEINSESIYVERTPFSSSSSKHSPSGTVA-----2777  
QY 2115 PGSRVSPENLVKSRGSRPGKSPERSHVSSEPYEPISPPQVPPVH--EKQDS 2164  
DB 2778 --ARVTPNY-----NPSRKSADSTARSQIP-TPVNNNTKKRDS 2817

## RESULT 8

US-08-289-548A-7  
Sequence 7, Application US/08289548A  
Patent No. 5648212  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERTIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Allegretti, LTD  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/289,548A  
FILING DATE: 12-AUG-1994



CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107,46943  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-289-548A-7

Query Match 2.98; Score 365; DB 1; Length 2842;  
Best Local Similarity 19.08; Pred. No. 6.6e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;  
Qy 347 PEIRKQREQEE-RFORVGORGAGLSATLARSEH-----EISEIIDGLSEQENNEK 395  
Db 980 PSIESYEDDESCKSYQYPADLAHKIHSANHMDDNDGELDTINYSLKYSDEQLNSGR 1039  
Qy 396 QMRQLSVTPMMFADAEORRVKFINNGLMEDPMKYKQDRQFMNVWTHD--KEIFKDKFI 453  
Db 1040 Q-----SPSONERWARPKHIIEDIEIKQEQSRQNSQSTTVVYVTESTDDKHL 1086  
Qy 454 QHPKNFGLIASYLERKSPVDCVLYLYLTKKENYKALVRNYKRRGRNQOIARPSOEK 513  
Db 1087 KPQHFH-----QOECVSP-----RSRGANGSTNRVGNHGINQVNSQLCOE- 1131  
Qy 514 VEEKEEDKAERTEKK--BEERKDEEKEDEKSKENTKEKDKIDGTABETEEREQAIPRG 571  
Db 1132 -DYEDDKPTNYSERYSEEQHEEERPTNYSIKYN-EERHVDOPIDYSLKYATDIPSS 1189  
Qy 572 R-----KTANSQRRKGRITRSMWNEAAAAS-----AAAAATEEPPPPPPP 614  
Db 1190 OKQSFSSKSSGSSQSSKTEHSSSSSENTSTPSSNAKRONLHPSSAQRSGQPKAAATCK 1249  
Qy 615 PEPTISTEPEV-----SRWTEEMEYAKGLVEGRN-----WAAIAKM 653  
Db 1250 VSSINQETIQTICVEDTPICFSRSSSSLSLSAEDIEGCGNTTQEAADSANTLQIAEIK 1309  
Qy 654 VGTKSE-----AQCKNFYENYKRRHLDNLLOQHK-----QKTSR 688  
Db 1310 IGTRSAEDPVSEVPASVQHPTKSSRLOGSSLSSESARHKAVERFSSGAKSPSKCAQTPK 1369  
Qy 689 KPREE-----RDVQCESVASTVSAQDEDIASNEENEDPESEVAVKPE--DSPEN 740  
Db 1370 SPPIHYVQETPLMFSRCTSV-SLSDSFESRSIASSVQPCSGMVSGIISPSDLPSDPGQ 1428  
Qy 741 A--TSRGWTEPAVELEPTETAPSTPSLAVPSKPADESVETQVNDISAEATAEQMDV 798  
Db 1429 TMPRSRSTPPPP-----PPQTAQTKREVKNKAPTAEKRESGPKQAAVNAVQ-RVQVLPDA 1484  
Qy 799 DQQHSABE-----GVC-----DPPATKADSDVDVEVR-----VPEN-HASKVEGDN 840  
Db 1485 DTLHFATESPTDGFSCSSLSALSLEPFIQK-----DVELRIMPVQENDNGNETESEQ 1540  
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQQTNAORPEFPQSDNDSSATCSADEVD----- 890  
Db 1541 PKSENQOEKAEKTIIDSEKDLL-----DDSD-----DDDIIELEECIIS 1580  
Qy 891 ---GEPEQRMPDMSKPSLNLPTGSIILVSSPLKPNPLD-----LP-----QLQHRAAVIP 938  
Db 1581 AMPTKSSRKAKKPAQATSKLPPPPVAR-----KPSQLPVYKLLPSQNRLOPKHVSFT 1633

Qy 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1634 GDDMPRYVEGTPIINFSTATISDLTIESPPNELAAGEVGRGGAQGEFEKRTIITEG 1693  
Qy 971 AL-----LEEQRQEQDILECRSSSTSPGOTSKSPNREWEVL-----1007  
Db 1694 RSTDEAOGGKTSSVTIPELDDNKAEGDILAEICINSAMPKGSHPKPRVKKIMDQVOQAS 1753  
Qy 1008 --QAPHQLTNLPGEVRLPTTRTRPPPLI-----PS 1039  
Db 1754 ASSAPNK---NQLDGKKKPTSPVKPIQNTYRTRVRKNADSKNNLNAERVFSDNKDS 1810  
Qy 1040 SKTTVAEKPSF-----IMGGSIS-----OGTPGYLTSHNQASVQTETPK 1080  
Db 1811 KKNLKNNSKDFNDKLPNNEDRVGRGFAFDSPHHYTIEGTP--YCFSRND-----1859  
Qy 1081 PSVGSISLGLPROQOESAKSATLPYIKQEEFSPRSONSOPEGLLYVRAQHEGVYVRGTAIG 1140  
Db 1860 -SLSLDFDDDDVDLSREKAELRKAKENKESEAKVTSHTE--LTSNQQA---NKTQAI 1913  
Qy 1141 EGSITRGTPTSKISVESIPSLRGSITOGTPALPOTGIPTALVK-----GS 1186  
Db 1914 KQPIRGOPKPILOQS-----TFPOSSKDIPORGAATDEKLNFAIENTPVCFSHNS 1967  
Qy 1187 ISRM-----PIEDSSPEKGRGEEAASKGHYIEGKSGHI-----1219  
Db 1968 LSSLSLDIDQENNNKENPIKETEPDSQGPSPKQASGYAPKSFHVEDTPVCFSRNSSL 2027  
Qy 1220 -LSDYN-----IKWAREGTRSPRTAHEISLKRSTVESVEGNIKQKMSMRSPVAPLE 1270  
Db 2028 SLSDSEDDLLQECISSAMPKKKPSRLKGDNEKHSRNMGGILGEDLTLDLKDQIRPDS 2087  
Qy 1271 GLICRALPRGSPHDLKERTVLGSIIMOGTPRATTESEFDGLKYPKOIKRESPIRAPEG 1330  
Db 2088 E---HGLSPDSENDWKAIOEGANSIVSSLHQAAAAAC-----LSRQASDSDLSLS 2139  
Qy 1331 AITKGKPYDGIITTIKEMGRSIEIPRODILTQESRKTPVQVQSTRPIEGSISOGTPIKF 1390  
Db 2140 GISLGSFP-----HLTPDQEEKPFTSNKGPRIK-----PGEKSTLETKKI 2180  
Qy 1391 DNNS-GSATAKHNVKSLITGP-----SKLSRGM-PPLEIVPENIKVGRKYEDVAGETV 1444  
Db 2181 ESEKIGKGGKVKVYKSLITKVRSENSEISGMQKPLQ---ANMPSISRG-----2226  
Qy 1445 RSRITSVSSGSPSVLRSLTHEAPKAQLSPGIYDDTSARRTPSVYQNTMSRGSPPMNRSTD 1504  
Db 2227 -----RTMIH-----IPGV-----RNSSSSTSPVSKKGPP 2251  
Qy 1505 VTIPPKNSTNHERKSTLTPT-QRESIPAK-SPVPQVDPVWVSHSPDPHHRGSTAGEVYWS 1562  
Db 2252 LKTPASKSPSEGOATTSIPRGAKPSVKSELSPVAROTSOIGSSKAPSRSGSR-----2304  
Qy 1563 HLPQLDPAFPFHRALDPAAYLFOQLSPTPGYPSQYQIYAMENTROTILNDIYTSQQ 1622  
Db 2305 -----DSTPSRAQQL-----SRP 2319  
Qy 1623 MOYNLRPDVA---RGLSPREQLPLPYPAIRGIIDLINMPPTILVPHGCTSTPMDRIT 1679  
Db 2320 IQSPGRNSISGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGSRMS 2364  
Qy 1680 YIPGTQITFFPPRYNSASMGPHPTHLAAAASAEERERERE-----KERERE 1727  
Db 2365 Y-----TSPGRMSQONLT--KQTGLSKNASSIPRESASKGLNMNGNGANKVELS 2416  
Qy 1728 RIAAASDLYLRPGSEOPGRPGSHGYVRSPSPSVRTQETMLQQRPSVFGQTNGTSVTPL 1787  
Db 2417 RMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS-----2448  
Qy 1788 DPTAQLRIMPLPAGGPSISQCLPASRYNTAADALAALVDAASAQMDVSKTKESKHEAA 1847  
Db 2449 -PTLR-RKLEESASFESLS---PSSRPASPTRSQAPTPVLSPLDMSLS--THSSVQAG 2502



Qy 1848 --RLEENLRSAVSEQQLEOKTLEVKRSVQCLYTSSAFFSGKQPKQPHSSVVYSAGK 1905  
Db 2503 WRKLPNLT--SPTIEYNGRPAKRHDIA-----SHSESPSRLPINRSGTWKREHSK 2552  
Qy 1906 DKGPPKRYEEELTRGKITITAN-----FIDVITRQIASOKDARERGSQSSDS 1957  
Db 2553 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEDEKHVNSISGTKQSKENQVSAGT 2610  
Qy 1958 SSSLSHRYE--TPSDATEVSPASSAPPQEKLOTYOPEVVKANAENDPTRQYEGPLHH 2016  
Db 2611 WRKIKENEFNTNSTQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2667  
Qy 2017 YRQOESPSPOQLPP-----SSQAGMGQVPRTHRLITLADHIC 2056  
Db 2668 PR---SGRPTGNTPPVIDSVSEKANPNIKDKQNAQKQNGVSGVPM--RTVGLENRLN 2722  
Qy 2057 QIIQTDFARNQVSSQTP--QQPPTTFQNSPVALYS--TPVRTKTNRNRPESQAOQSVIHQR 2114  
Db 2723 SFIQVADAPQKGTRIKPKQNNPVPVSETNESSIVERTPFSSSSSKHSSPSGTVV----- 2777  
Qy 2115 PGRSVSPENLVKSGRSGRPGKSPERSHVSSEPEYPIPPQVPPVH--EKQDS 2164  
Db 2778 --ARVTFPNY-----NPSPRKSSADTSARPSOIP--TPVNNNTKKRDS 2817

RESULT 9  
US-08-452-654-7  
Sequence 7, Application US/08452654  
Patent No. 5691454  
GENERAL INFORMATION:  
APPLICANT: ALBERTSEN, HANS  
APPLICANT: ANAND, RAKESH  
APPLICANT: CARLSON, MARY  
APPLICANT: GRODEN, JOANNA  
APPLICANT: HEDGE, PHILIP J.  
APPLICANT: JOSLYN, GEOFF  
APPLICANT: KINZLER, KENNETH  
APPLICANT: MARKHAM, ALEXANDER F.  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: THLIVERIS, ANDREW  
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,654  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.035574  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2842 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: Homo sapiens  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-452-654-7

Query Match 2.9%; Score 365; DB 1; Length 2842;  
Best Local Similarity 19.0%; Pred. No. 6.6e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQEQOE-RPQRQRCAGLSATIRSEH-----EISEIIDLSQEENNEK 395  
Db 980 PSIESYEDDESKEFCSGYQPADLAHLIHSANHMDNDGELDTPTINYSKYSQSDOLNSGR 1039  
Qy 396 QMRQLSVIPPMFADAEQRRVKFINNGLMEDPMKVYKDRQFMNVTQHE--KEIFKDKFI 453  
Db 1040 Q-----SPQNERWARPKHIIIEIKQSEQRQSRNQSTYTPVYTESTDDKHL 1086  
Qy 454 QHPKNFGLIASYLERKSVDPCLVYYITKKKNYKALVRRNYGKRGRNQOIARPSQEEK 513  
Db 1087 KFQPHFG-----QOECVSPY-----RSRGANGSETNRVSGNHGINQVQSLOQE- 1131  
Qy 514 VEEKEEDKAEKTEK--EEKKDEEKDEKEDSKENTKEKDKIDGTAEETEEREQATPRG 571  
Db 1132 -DDYEDDKPTNYSRYEEEOHEEERPTNYSIKYN--EERHVDOPIDYSUKYATDIPSS 1189  
Qy 572 R-----KTANSQGRKRGRITRGMTNEAAAAA-----AAAAAATEEPPLPPP 614  
Db 1190 QKQSFYSKSSSGSSGSKTEHSSSENSTPSSNAKRNQNLHPSSAQSRSGQPKATCK 1249  
Qy 615 PEPITSETPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
Db 1250 VSSINQETIOTIYCYVEDTPICFSRCSLSLSAEDICGNOTQOADSANTLQIAEIK 1309  
Qy 654 VGTKSE-----AQCKNFYNYKRRHNLNLOQHK-----QKTSR 688  
Db 1310 IGRTSAEDPVSEPAVSQHPRTKSRLOGSSLSSESAHKAHKAHKAHKAHKAHKAHKAH 1369  
Qy 689 KPREE-----RDVSOCEVASTVSAQEDIEDIEASNEENPEDESEAVKPSF--DSPN 740  
Db 1370 SPPEHYVOETPLMFSCRSTV--SLDSFESRSIASSVQPCSGMVSGIISPSDLPSPGQ 1428  
Qy 741 A--TSRGNTPAVELEPTTETAPSTPSLAVSTPKPAEDSVETQVNDISIAETAEQMDV 798  
Db 1429 TMPFSRSKTPPP--PFQTAQTKREVPKNKAPTAEKRESGPKQAANAQV--RVQVLPDA 1484  
Qy 799 DQOEHSAEE--GSVC-----DPPPATKADSDVDVEVR--VPEN--HASKVEGDN 840  
Db 1485 DTLHFATESTPDGFSGSSLSLSALSDLEPFIQK----DVELRIMPVOENDNGNETESE 1540  
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQAQINQARPEPQSDNDSATCSADEVD----- 890  
Db 1541 PKESNEQNEKEAKTIDSEKDLL-----DQSD-----DDDIEILEECIIS 1580  
Qy 891 ---CEPERQRMFMDSKPSSLNPTGSLVSSPLKPNPLD----LP-----QLQRAAVIP 938  
Db 1581 AMPTKSSRKAQKPAQATASKLPPVVAR-----KPSQLPVTKLLPSQNLQPKQHVSTP 1633  
Qy 939 ---PMVSC--TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1634 GDDMPRYCVEGTPINFSTATSLDITESPPELAGEVGRGQAQSGEFKRDRTIPTEG 1693  
Qy 971 AL-----LEQORQEQIDLECRSSSTSPCGTSKSPNREWEVL----- 1007  
Db 1694 RSTDEAOGGKTSSTVTPELDNDKAEEDGILAEICINSAMPKSKHKKPRVKKIMDOVQOAS 1753  
Qy 1008 --QPAPHQITNLPEGVRLPTTRTPPPPLI-----PS 1039  
Db 1754 ASSSAPNK---NQLDGKKKKTPSPVKPIQNTYRTRVYRNKADSKNNLNAERFVSDNKDS 1810



QY 1040 SKTTVASEKPSF-----IMGSIS-----QGTPTGYLTSHNOASYTQETPK 1080  
Db 1811 KKONLKNNSKDFNDKLPNNEDVRGSPAFDPSHHVTPTEGTP--YCFSRND-----1859  
QY 1081 PSVGSISGLRQPBESAKSAILPYIKQEFSPRSONSOPEGLLVRAQHEGVVRGTAGAIQ 1140  
Db 1860 -SLSLDFDDDDVDLREKAKELRAKENKESEAKVTSHTS--LTSNQOSA---NKTQAI 1913  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQCTPALPOTGITEALVK-----GS 1186  
Db 1914 KOPINRGQPKPILOKS-----TFQSKDIPDRGAATDEKLQNFALIENTPVCFSHNS 1967  
QY 1187 ISRM-----PIEDSSPEKGREAAKSHVIEGKSGHI-----1219  
Db 1968 LSSLSIDIDENNKENNEPIKETEPDSDGEPFKQASGYAPKSFHVEDTPVCFSRNSSL 2027  
QY 1220 -LSYDN-----IKNARETRSPRTAHEISLKRSEVSEGNKQKQMSRESVPVAPLE 1270  
Db 2028 SLSIDSEDDLQECISSAMPKKKPKSRLLKGDNEKHSRNMGGILGEDLTDLKDIQRPDS 2087  
QY 1271 GLICRALPRGSPHDLKERTVLGSGIMOGTTRATTESPEDGLKYPKQIKRESPIRAPEG 1330  
Db 2088 E---HGLSPDSEDFWKAIQEGANSVSSLLHQAAC-----LSRQASSDSLSLSKS 2139  
QY 1331 AITGKPYDGIITIKEMGRSIEIPRQDILTQESRKTPEVVQSTRPIIEGSIQGTPIKF 1390  
Db 2140 GISLGSPE-----HLPTDQEKPTTSNKGPRILK-----PGEKSTLETKKI 2180  
QY 1391 DNNS-GQSAIKHNKSLTGP-----SKLSRGM-PPEIVPENIKVVERGYEDYKAGETV 1444  
Db 2181 ESESGIKRGGKVKYLSLTGKVRNSSEISGOMKQPLQ---ANMPSISRG-----2226  
QY 1445 RSRHTSVVSGSPVLSRLTLEAPRAQLSPGIYDDTSARRTPVSYQNTMSRSGPMNRTSD 1504  
Db 2227 -----RTMIH-----IPGV-----RNSSSSTSPVSKKGP 2251  
QY 1505 VTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPQVDVPSVSHSPDFPHRGSTAGEVYWS 1562  
Db 2252 LKTPASKSPSEGOYATTSFPRGAKSVKSELSFVARQTSIQIGSSKAPSRSGSR-----2304  
QY 1563 HLPQLDPAMPFHRALDPAALYLFQRLSTPTGVPYQYLIYAMENTRQTILNDYITSQQ 1622  
Db 2305 -----DSTPSRPAQOPL-----SRP 2319  
QY 1623 MOVNLRPDVA--RGLSPREQPLPYPATRGIDLTNMPPTILVPHPGTSTPPMDRIT 1679  
Db 2320 IQSPGRNISGRNGISPPNK-----LSOLPRT-SSPSTASTKSSGSGRMS 2364  
QY 1680 YIPGTQITFPPRPNYSMSMCPHTHLAAAASAERERERE-----KERERE 1727  
Db 2365 Y-----TSPGRQMSQOQLT--KOTGLSKNASTIPRESASKGLNOMNGNANKVELS 2416  
QY 1728 RIAAASDLYLRPGEQGRGSGHYVRSPSPSVRTQETMLQQRPSVFGTNGTNGSVITPL 1787  
Db 2417 RMSSTKSS-----GSESD-----RSERPVLVRQSTFIKEAPS-----2448  
QY 1788 DPTAQLRIMPLPAGSPSTISQGLPASRYNTAADALALVDAASAPQMDVSKTESKHEAA 1847  
Db 2449 -PTLR-RKLESASFESUS---PSRPASTRSQAPQTPVLSPLDMSLS--THSSVQAGG 2502  
QY 1848 ---RLEENLRSAVSEQQOQLEQTKLEKRSVQCLYTSSAPFSGKQPKPHSSVYVSEAGK 1905  
Db 2503 WRKLPPNL---SPTIEYNDGRPAKRHDIAF-----SHSESPSRPLPINRSGTWKREHSK 2552  
QY 1906 DKGPPPKRYEEELTRGKTTITAN-----FIDVITRQIASDKDARERGSQSDS 1957  
Db 2553 HSSSLP--RVSTWRTGSSSILSASSESESEKAKSEKHNVSISGTRKQKNOVSAKGT 2610  
QY 1958 SSSLSSSHRYE--TPSDATEIVSPASSAPPQEQKLTQYQEVVKAQNAENDPTQVEGLPHH 2016  
Db 2611 WRKIKENEFTNTSQTIVSSGATNGAESKTLIYQMAFVSK---TEDVWVRIEDCPINN 2667

QY 2017 YRQOQSPSPQOQLPP-----SSQAEQMGOVPRTHRLITLADHIC 2056  
Db 2668 PR---SGRSTGNTPPVIDSVSEKANPNIKSDKNQAKQNGVSGVPM--RTVGLNRLN 2722  
QY 2057 QIITQDFARNQVSSQTP-QQPPTTFQNSPALVS-TPVTKTNSRYSPEQASQSVHHOR 2114  
Db 2723 SFIOVDAPDQKGTIKPQNNPVPVSETNESSIVERTPFSSSSSKHSPSGTVA-----2777  
QY 2115 PGRSVSPENLVKSRGSRPKSPERSHVSSEPEPIPSPPQVPPVH--EKQDS 2164  
Db 2778 --ARVTFNY-----NPSPRKSSADSTSARPSQIP-TPVNNNTKKRDS 2817

## RESULT 10

US-08-452-655B-2  
; Sequence 2, Application US/08452655B  
; Patent No. 5783666  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,655B  
; FILING DATE: 25-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/289,548  
; FILING DATE: 12-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107,49964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-452-655B-2

Query Match 2.9%; Score 365; DB 1; Length 2843;  
Best Local Similarity 19.0%; Pred. No. 6.6e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;



QY 347 PEIKOREQOE-RFORVGORGAGLSATIARSEH-----EISEITDGLSEONNEK 395  
Db 981 PSIESYSEDDSKSCYQYADLAHKHSANHMDDNDGELDTPLNYSLKYSDEDLNNGR 1040  
QY 336 OMROLSVIPPMFMDAORRVKFINNMGLMEDPMKYKIDROPFMYVWTDHE--KEIFKDFI 453  
Db 1041 Q-----SPSQNERWARPKHIIIEDEIKOSEQORQNOQSTTPVVTTESTDDKHL 1087  
QY 454 QHPKNFGLIASYLERKSPVDCVLYLYLTKKNENYKALVRNRYKRRGRNOQIARPSQEK 513  
Db 1088 KFOPHF-----QOECVSP-----RSRGANGSTNRVGSNHGINQVQSLOQE- 1132  
QY 514 VEEKEEDKAETK--BEEKDDEEKDEKEDSKENTKEDKIDGTABETEEREQATPRG 571  
Db 1133 -DDYEDDKPTNYSERYSEEOHEEERTNYSIKYN-EERHVDOPIDYSUKYATDIPSS 1190  
QY 572 R-----KTANSQRRKGRITRSMNNAAS-----AAAAATEEPPPPPPP 614  
Db 1191 OKQSFSSKSSGSSGSKTEHMSSESENTSTPSSNAKRONQLHPSSAQSRGQPKAAATCK 1250  
QY 615 PEPISSTEPVE-----SRWTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
Db 1251 VSSINQETIQTICVEDTPICFRSGSSLSLSAEDEICGNOTTOEADSAANTLQIAETKEK 1310  
QY 654 VGTKSE-----AQCKNFYNYKRRHNLNLLQOHK-----QKTSR 688  
Db 1311 IGTRSAEDPVSEPAVSOHPRTKSSRLOGSSLSSESARHKAVERFSSGAKSPSKSQAOTPK 1370  
QY 689 KPREE-----RDVSOQESVASTVSAQDEEDIASNEEENEDSEVEAVKPYSE--DSPEN 740  
Db 1371 SPPEHYVOETPLMFSTRCTSV--SSLDSFESRSIASVSEPCSGMVSGIISPSDLPSDQ 1429  
QY 741 A--TSRGNTAPEVALEPTETAPSTPSLAVPSTKPADESVETQVNDISIAETAQMDV 798  
Db 1430 TMPRSRSTPPP--PPQTAQTKREVFNKAPTAKEKRESGPKQAANVAQV-RVQVLPDA 1495  
QY 799 DOQHSABE-----GSVC-----DPPATKADSDVDVEYR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHFATFESTPDGFCSSLSALSALDLEPFIQK-----DVELRIMPPVQVONGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPDEDLVAQINAQRPQSDNDSATCSADEVD-----890  
Db 1542 PKSENQEKAEKTIIDSEKDLL-----DDSD-----DDDEILEECIIS 1581  
QY 891 ---GEPERQRMFPDMSPLNPTGSLVSSPLKPNPLD---LP-----OLOHRAAVIP 938  
Db 1582 AMPYKSSRKAAPKPAQATASKLPPVAR-----KPSQLPVYKLLPSQNRLOPKQHVSETP 1634  
QY 939 ---PMVSC-----TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRVYCVGTPINFSTATSLDLATIESPPNELAAGEVGRGAQSGEFEKRDITIPEG 1694  
QY 971 AL-----LEQORQOEIDLECRSSTSPCGTSKSPNREWEVL-----1007  
Db 1695 RSTDEAOGKTSVTIPELDDNKAEBGDILAEICNSAMPKKGSHKPFVRKKIMDQVQOAS 1754  
QY 1008 --QAPAHOLITNLEPGRVLRTPTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NOLDGKKKFTSPVKPIQNTYETRVKRNKADSKNNLNAREVFSNDKDS 1811  
QY 1040 SKTTVASEKPF-----IMGGSIS-----QGTPTGYLTSHNOASYTQETPK 1080  
Db 1812 KQNLKNNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTP--YCFSRND-----1860  
QY 1081 PSVGSISGLPROQESAKSATLPYKQEFSPRSQNSOPEGLLVRAQHEGVVGRTAGAIQ 1140  
Db 1861 -SLSLDFDDDDVDSREKAEKELRAKENKESEAKVTSHTE--LTSNOOSA---NKTOAIA 1914  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGOPRPILOQS-----TFPOSSKDIPIRGAATDEKQLONFAIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSSPEKGREEAASKGHVIEGKSGHI-----1219

Db 1969 LSSLSIDDOENNNKENEFIKETEPDSCQEFKSKPOASGYAPKSFHVEDTPVCFSRNSSL 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKSYSVESVEGNIKQGMRESYPASLE 1270  
Db 2029 SLSIDSEDDLQECISSAPKKKKPSRLKGDNEKHSRNMGGILGEDTLDLKDIQRPDS 2088  
QY 1271 GLICRAIPRGSPHSDLKERTVLSGSIMOGTTPRATTESEFDGLKYPKQIKRESPPIRAEG 1330  
Db 2089 E---HGLSPDSENFENDKAIQEGANSIVSSLHQAASAAAC-----LSRQASSDSDLSLS 2140  
QY 1331 AITKGPVGDITTIKEMGRSIEHPRDILQESKTEPEVQVSTRPIIEGTSQGTPTKF 1390  
Db 2141 GISLSP-----HLTPDOBEKFTSNKGRILK-----PGEKSTILETKI 2181  
QY 1391 DNNS-GOGAIKHNVKSLITGP-----SKLSRGM-PLEITVPENIKVVERGYEDVKAQETV 1444  
Db 2182 ESEKGIKGGKVKVYKSLITGKVRNSSETSGOMKOPLO---ANMPSISRG-----2227  
QY 1445 RSRHTSVVSSGSPSVLRSTLHEAPKAQLSPGIYDDTSARRTPVSQNTMRSRSPMMNRISD 1504  
Db 2228 -----RTMIH-----IPGV-----RNSSTSPSVSKGPP 2252  
QY 1505 VTIPPKNSTNHERKSTLTPT-QRESIPAK-SPVPCVDVPVWSHSPFDPHHRSSTAGEVYWS 1562  
Db 2253 LKTASKSPSEGOATTSPRGAKPSVKSELSFVARQTSQIGSSKAPSRSGSR-----2305  
QY 1563 HLTQLODAMPFHRALDPAANAAYLFOROLSTPGYPSQYLYAMENTROTILNDYITSOQ 1622  
Db 2306 -----DSTSRPAQOPL-----SRP 2320  
QY 1623 MQVNLRPDVA--RGLSPREQPLGYPATRGIIIDLTNMPPTILVPHPGTSTPMDRIT 1679  
Db 2321 IQSGRNSISGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGGKMS 2365  
QY 1680 YIPQTIQTFPPRPNYSASMPGHPHTLAAAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRQMSQOQLT--KOTGLSKNASSIPRSESASKGLNMNGNANKKVELS 2417  
QY 1728 RIAAASDLYLRPSEOPGRPGSHGYRSPSPSVRTQETMLQORPSVPOGTINGSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSERPVLRVROSTFKEAPS-----2449  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADAALAALVDAASAPQMDVSKTESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFESLS---PSRRPASPTRSQAQTPVLSPLPDMXLS-THSSVQAGG 2503  
QY 1848 --RLEENLRSSAAVSEOOQLEQKTEVEKRSVQCLYTSSAPPSGKPOPHSSVYVSEAGK 1905  
Db 2504 WRKLPPNL---SPTIEYNDGRPAKRHDIAI-----SHSESPSRLPINRSQWKRHSK 2553  
QY 1906 DKGPPPKSRYEELLRGKTITITAAAN-----FIDVITRQIASDKDARERSQSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASESESEKAKSEDEKHVNSISGTSKQENQVSAKGT 2611  
QY 1958 SSSLSSSHRYE--TPSDAIEVISPASSPAPQEKLOTQOEVVYKANOENDPTROVEGLUHH 2016  
Db 2612 WRKIKENEFTNSTQTSVSGATNGAESKTLIYQMAPAVSK---TEDVWTRIEDCPINN 2668  
QY 2017 YRPOQESPSPOOQLPP-----SSQAEGMGQVPRTHRLITLADHIC 2056  
Db 2669 PR---SGRSPGTGNTPPVIDSVEKANPNIKSKDNQAKONVGNNGVPM--RTVGLENRLN 2723  
QY 2057 QILITQDFARNQVSSQTP-QQPPTSTFQNSPSALYS-TPVTKTNRNYSPEQASQVHHQR 2114  
Db 2724 SFIQVADAPDQKGTIKPQNNPVPVSETNESSIVERTPFSSSSSKSHSSPGTVA-----2778  
QY 2115 PGSRVSPENLVKSRGSRGKSPERSHVSSEPEYPISPQVPPVH--EKQDS 2164  
Db 2779 --ARVTPFNY-----NPSPKSSADSTARSQIP-TPVNNNTYKKRDS 2818

RESULT 11



US-08-452-655B-7

; Sequence 7, Application US/08452655B

; Patent No. 5783666

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

; APPLICANT: CARLSON, MARY

; APPLICANT: GRODEN, JOANNA

; APPLICANT: HEDGE, PHILIP J.

; APPLICANT: JOSLYN, GEOFF

; APPLICANT: KINZLER, KENNETH

; APPLICANT: MARKHAM, ALEXANDER F.

; APPLICANT: NAKAMURA, YUSUKE

; APPLICANT: THLIVERIS, ANDREW

; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

; NUMBER OF SEQUENCES: 102

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner &amp; Witcoff, Ltd.

; STREET: 1001 G Street, NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20001-4598

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/452.655B

; FILING DATE: 25-MAY-1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/289,548

; FILING DATE: 12-AUG-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/741,940

; FILING DATE: 08-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Kagan, Sarah A.

; REGISTRATION NUMBER: 32,141

; REFERENCE/DOCKET NUMBER: 1107.49964

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-508-9100

; TELEFAX: 202-508-9299

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2843 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHEICAL: YES

; ANTI-SENSE: NO

; US-08-452-655B-7

Query Match 2.98; Score 365; DB 1; Length 2843;

Best Local Similarity 19.08; Pred. No. 6.6e-14;

Matches 409; Conservative 289; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQEQB-RFQRVGQAGLSATIRSEH-----BISETIDGLSQENNEK 395

Db 981 PSIESSEDESKFCYQYADLAHKSANHMDNDGELDTPIYSLKYSDQLNSGR 1040

Qy 396 QMRQLSVIPMMFAEORRVFINNGLMEDPMKYKDRQPMNVWDHE--KEIFKQKFI 453

Db 1041 Q-----SPSQNERWARPKHIEDEIKQSEQRQSRNQSTTPYVYTESTDDKHL 1087

Qy 454 QHPKNFGLIASLYERKSPVDCVLYLTTKKNYKALVRRNYGRRGRNQOIARPSOEK 513

Db 1088 KFQPHFG-----QOECVSPY-----RSRGANGSETNRVGNHGINQVNSLQOE- 1132

Qy 514 VEEKEDKAETK--EEEKDDEEKDEKSDKENTKEKIDKTAEETEEREQATPRG 571  
Db 1133 -DYEDDKPTNYSERYSEEEQHEEERTNYSIKYN-EERHVDQPIDYSKYATDIPSS 1190  
Qy 572 R-----KTANSQRRKGRITRSMTNEAAAAS-----AAAAATEEPPLPPPP 614  
Db 1191 QKQSFSEKSSGSSGSSKTEHSSSSSENTSTPSSNAKRNQNLHPSSAQSRGQPOKAATCK 1250  
Qy 615 PEPITSTEPVET-----SRWTEEMEYAKKGLVEHGRN-----WAAAKM 653  
Db 1251 VSSINQETIQTICVEDPPICFSRSSLSLSAEDEICGNTQTQEADSANTLQIAETKEK 1310  
Qy 654 VGTKSE-----AOCKNFYENYKRRHNLNLOQHK-----OQTSR 688  
Db 1311 IGTRSAEDPVSEPAVSQHPKTKSRLOGSSLSSESARHKAFFESGAKSPSKSQAOTPK 1370  
Qy 689 KPREE-----RVSOCESVASTVSAQDEDEIEASNEEENPEDESEAVKPSE--DSPEN 740  
Db 1371 SPPEHYVOETPLMFSTRCTSV--SSLDSESRSIASSVQSEPCSGMVSGIISPSDLPSPGQ 1429  
Qy 741 A--TSRGNTEPAVELEPTTETAPSTPSLAVPSTKPADESVETQVNDISAEAEQMDV 798  
Db 1430 TMPPSRSKTPPP-----PPOTAQTKREVKNKAPTAEKRESGKQAAVNAVO--RVQVLPDA 1485  
Qy 799 DQOEHSAEE-----GSVC-----DPPATKADSVDEVR-----VPEN-HASKVEGDN 840  
Db 1486 DTLHFATSTPDGFCSSSSLSALSDEPFIQK----DVELRIMPVQENDNGNETESEQ 1541  
Qy 841 TKE--RDLDRASEKVEPRDELVVAQOINAOPEPQSDNDSSATCSADEVD-----890  
Db 1542 PKESNEQKEAETIDSEKDL-----DDSD-----DDOIELEECIIS 1581  
Qy 891 ---GEPEQRQMFMDSKPSLLNPTGSLVSSPLKPNLD---LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSRKAKKPAQATASKLPPPVAR-----KPSQLPVYKLLPSQNRQPQKHVSFT 1634  
Qy 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPRYCVGEGTPIFSTATSLDITIESPPNELAAGEVGRGAQGEFEKRTITFEG 1694  
Qy 971 AL-----LEEQRQEQOIDLERSSTSPGCTSKSPNREVL-----1007  
Db 1695 RSTDEAOGKTSVVTIPELDDNKAEEGDIILAEICINSAMPKSKHKPRVKKIMDOVOQAS 1754  
Qy 1008 --QPAPHOLITNLEPGVRLPTTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NQLBGKKKPTSPVKPIQONTYRTRVRKNADSKNNLNAERVSFSDNKS 1811  
Qy 1040 SKTTVASEKPSF-----IMGSGIS-----QGTPTGTYLTSHNOASYTOETPK 1080  
Db 1812 KKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTP--YCFSRND-----1860  
Qy 1081 PSVGSISLGLPROESAKSATLPIYKOEFPSPRSONSOPEGLIVRAOHEGVVRVTAQAIQ 1140  
Db 1861 -SLSLDFFDDDDVLSREKAEELRAKENKESEAKVTSHTE--LTSNQOSA---NKTQAI 1914  
Qy 1141 EGSITRGTPTSKISVESIPSLRGSITOGTPALPOTGTPTEALVK-----GS 1186  
Db 1915 KQPINRQOPPILOKQS-----TFPOSSKDIPDGAATDEKLNFAIENTPVCFSHNS 1968  
Qy 1187 ISRM-----PIEDSSPEKGREAAKSGHVIYEGKSGHI-----1219  
Db 1969 LSSLSIDQENNNKENEPKETETPPDSQGEPSKPAQSGYAPKSPHVEDTPVCFSRNSLS 2028  
Qy 1220 -LSVDN-----IKVAREGTRSPRTAHSLSKRSVESVEGNIKQGMSRESVPSAPLE 1270  
Db 2029 SLSIDSEDDLQECISSAMPKKKPSRLKGDNEKHSRNMGMGILGEDLTLDLKDQIQRPS 2088  
Qy 1271 GLICRALPRGSPHDLKERTVLSGSIQOGTPRATTESFEDGLKVPKQIKRESPPIRAFEG 1330  
Db 2089 E---HGLSPSENFEDWAKIQEAGANSIVSSLHQAAAAAC-----LSRQASSDSDSLSLS 2140



Qy 1331 AITKGPYDITIKEMGRSIEIPRODILQESKRTEPVVQSTRPIIEGSIQCTPIKF 1390  
Db 2141 GISLGSPE-----HLTPDQEKFPSTNKGPRILK-----PGEKSTLETCKI 2181  
Qy 1391 DNNS-GQSAIKHNKSLITGP-----SKLSRCM-PPEIVPNIKIVVERGKDYKAGETV 1444  
Db 2182 ESEKGIKGGKVKVSLTIGVRSNSETSGOMKQPLQ---ANMPSISRG-----2227  
Qy 1445 RSRHTSVVSSGSPVSLRSTLHEAPKAQLSPGIYDDTSARRTPVSVQNTMRSGSPMMNRSTD 1504  
Db 2228 -----RTMIH-----IPGV-----RNSSTSPVSKGKPP 2252  
Qy 1505 VTIIPNKSTNHERKSTLTPT-QRESIPAK-SPVPGVDVVSHPDHPHHRGSTAGEVYWS 1562  
Db 2253 LKTPASKSPSEGTATTSPRCAKSPVKSELSVPARQTSQIGSSKAPSRSGSR-----2305  
Qy 1563 HLPQLDPAMPFHRALDPAALYLFQROLSPTPGVPQYQLYAMENTQTLINDYITISQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320  
Qy 1623 MOVNLRPDVA---RGLSPREOPLGLPYPATRGIIDLTNMPPTILVPHPGTSTPMDRIT 1679  
Db 2321 IQSPGRNISGRNGISPKN-----LSQLPT-SSPSTASTKSSGSGRWS 2365  
Qy 1680 YIPQTIPTPPRPNYSASMSGPHPTLAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQOQLT---KOTGLSKNASSIPRSASAKGLNOMNGANKKVELS 2417  
Qy 1728 RIAAASSDLYLRPGSEQPRGSHGYSRPSVYRTQETMLQORPSVFOQTNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSED-----RSERPVLVROSTTFKEAPS-----2449  
Qy 1788 DPTAQLRIMPLPAGPSTISQGLPASRYTAADALAALYDAASAQMDVSKTKESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFELS---PSSRPASPTRSQATPVLSPLDMSLS-THSSVQAGG 2503  
Qy 1848 --RLEENLRSAVSEOOQLEQKTEVEKRSVQCLYTSSAFFPKGPKOPHSPVSVSEAGK 1905  
Db 2504 WRKLPPNL---SPTIEYNDGRPAKRHDIAI-----SHSESPRLPTNRSQTKWKRHSK 2553  
Qy 1906 DKGPPPKSRYEEELTRGKTTITAN-----FIDVITRQIASDKDARERGSSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILGASSESEKAKSEDEKHVNSISGTKSKENQVSAKGT 2611  
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPQEKLTQYQEVVKANQAEKNDPTRQYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTQTSVSGAINGAESKTLIQMAFAVSK---TEDVWVRIEDCPINN 2668  
Qy 2017 YRPOOESPPOQLPP-----SSQAEGMGQVPRTHRLITILADHIC 2056  
Db 2669 PR--SGRSPGTNTPPVIDSVSEKANPNIKDSKONQAKQNGVNGSVPM--RTVGLNRLN 2723  
Qy 2057 QIITQDFARNQVSSQTP-QQPTSTFQNSPALYS-TPVRKTKNRYSPESQAOVSHQR 2114  
Db 2724 SFIOVDAPDQKTEIKGQNNPVPVSETFNESSIVERTPFSSSSSKHSSPSGTVA-----2778  
Qy 2115 PGSRVSPNLVDKSRGPRGSKPSKSHVSPSEYEPISPPQVPPVH--EKQDS 2164  
Db 2779 --ARVTPNY-----NFSPPKSSADSTSARPSQIP-TPVANNTKKRDS 2818

RESULT 12  
US-08-450-582-2  
; Sequence 2, Application US/08450582  
; Patent No. 6114124  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSLYN, GEOFF  
; APPLICANT: KINZLER, KENNETH

; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001-4598  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/450,582  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/452,655  
; FILING DATE: 25-MAY-1995  
; APPLICATION NUMBER: US 08/289,548  
; FILING DATE: 12-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/741,940  
; FILING DATE: 08-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A.  
; REGISTRATION NUMBER: 32,141  
; REFERENCE/DOCKET NUMBER: 1107.49964  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2843 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-450-582-2

Query Match 2.9%; Score 365; DB 3; Length 2843;  
Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

Qy 347 PEIRKQEQQE-RFQVVGQRGAGLSATIRSEH-----EISEIDGLSEQENNEK 395  
Db 981 PSIESYDEDESKFCYQYPADLAHKIHSANHMDNDGELDTPINYSLKYSDEQLNSGR 1040  
Qy 396 QMRQLSVIPPMFDAEQRRVKFINNGLMEDPMKYKDRQFMVNTDHE--KEIFKDKFI 453  
Db 1041 Q-----SPSQNERWARPKHIEDEIKQSEQRQSRNQSTTYPVYTESTDDKHL 1087  
Qy 454 QHPKNFGLIASYLERKSVDPDCLVLYLTKKNENTKALVRRNYGKRGRNQIARPSQEEK 513  
Db 1088 KFQPHFG-----QOECVSPY-----RSRGANGSETNRVGSNHGINQVSLCOE- 1132  
Qy 514 VEEKEEDKAEKTEKK--EEEKKDEEKDEKEDSKENTKDKIDGTAETETEREQATPRG 571  
Db 1133 -DDYEDDKPTWYSERYSEEEQHEEERTNYSIKYN-EKRRHVDPQIDYSLUKYATDIPSS 1190  
Qy 572 R-----KTANSQGRKGRITRSMTEAAAAA-----AAAAATEEPPLPPPP 614  
Db 1191 QKQSFSEKSSGSSGSKTEHSSSENSTPSSNAKRONQLHPSSAQSRSGQPQKAATCK 1250  
Qy 615 PEPITSTEPVET-----SRWTEEMEVAKKGLVEHGRN-----NAAIAKM 653  
Db 1251 VSSINQETIQYCVEDTPICFSRCSLSLSSAEIDGICNQTQEQADSANTLQIAEIKEK 1310



QY 654 VGTKE-----AOCKNFYKRRHNDLLOQHK-----QKTSR 688  
Db 1311 IGTRAEAPVSEVPAVSQHPRTKSRGLSSLSSESARHKAVERSSGAKSPKSGAQTPK 1370  
QY 689 KPRE-----RDVSCSEVASTVSAQDEIDEASNEBENEDSEVEAVKPSE--DSPEN 740  
Db 1371 SPPEHYQETPLMFSRCTSV--SSLDSFESRSIASSVQSEPCSGMVGIIISPSDLDPSPQ 1429  
QY 741 A--TSRGMTEPAVELEPTETAPSTPSLAVPSTKPADESVETQVNDISIAETAEDMV 798  
Db 1430 TMPSPRSKTPPP--PPOTAQTKREVPRNKAPTAEKRESGPKQAANVAQV--RVQVLPDA 1485  
QY 799 DQEHSAEE-----GSVC-----DPPATKADSVDEVYR--VPEN--HASKVEGDN 840  
Db 1486 DTLHFATESTPDGFCSSSLSALSDEFFIQ-----DVELIMPPVOENONGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPDEDLVVAQINAQRPQSDNDSSATCSADEYD-----890  
Db 1542 PKESNEQKEAETIDSEKOLL-----DDSD-----DDDIIELEECIIS 1581  
QY 891 ---GEPEQRMPDMSKSLNPTGSIIVSPLKPNPLD---LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSSKAKKPAQATSKLPPVAR-----KPSQLPVYKLLPSQNRLOPKQHVSTP 1634  
QY 939 ---PMVSC---TPCNIPIGTVPVGYAL-----YORHIKAMHES 970  
Db 1635 GDDMPVYCVGCTPINFSTATSLDLTIESPPNELAAGEVGRGAQSGEFKRTIPTEG 1694  
QY 971 AL-----LEBQRQEQIDLECRSSTSPCGTSKSPNREVL-----1007  
Db 1695 RSTDEAOGKTSSTVPELDDNKAEEGDILAEICINSAMPKSKHKPRVKKIMDQVOQAS 1754  
QY 1008 --QAPHOLIINLEGVRLPTRTRPPPLI-----PS 1039  
Db 1755 ASSAPNK---NOLDGKKKTPSVKPIQNTYRVRKNADSKNNLNAERFVSDNKDS 1811  
QY 1040 SKTVASEKPSF-----IMGGSIS-----QGTGYLTSHNQASYTQETPK 1080  
Db 1812 KKONLKNNSKDFNDKLPNNEDVRGSAFADSPHHYTPIEGTP--YCFSRND-----1860  
QY 1081 PSVGSISGLPQOESAKSATLPYKQEEFSPRSONSOPEGLLVRAQHEGVVGTAGAIQ 1140  
Db 1861 --SLSLDDDDVDLSREKAEKRAKENKESEAKVTSHT--LTSNQOSA---NKTOAIA 1914  
QY 1141 EGSITRGPTSKISVESIPSLRGSITQCTPALPOTGIPTALVK-----GS 1186  
Db 1915 KQPINRGQPKILQKS-----TFQSSKDIIPDGAATDEKLNFAIENTPVCFESHNS 1968  
QY 1187 ISRM-----PIEDSSPEKGREAAKSHVYIEGKSGHI-----1219  
Db 1969 LSSLSDIDQENNNKENEPKETEPDSDGEPKPOAGYAPKSPHVEDTPVCFSRNSSL 2028  
QY 1220 --LSYDN-----IKNARETRSPTRAHEISLKRSESVEGNIKQGMRESVPSPAPLE 1270  
Db 2029 SLSIDSEDLLOECISSAMKKKPSRLKGNKHSRPNMGILGEDITLDLKDIOQRPDS 2088  
QY 1271 GLICRALPRGSPHDLKERTVLSGIMOGTPRATESPEDGLKVPKQIKRESPPITAFEG 1330  
Db 2089 E--HGLSPDSENFDAKIQAGANSIVSSLHQAAAAC-----LSRQASSDSLSLSKS 2140  
QY 1331 AITGKPKYDGIITIKEMGRSITHEIPRODILQESRKTPVQVSTRPIEGSISQGTPIKF 1390  
Db 2141 GISLGSPP-----HLTPQOEKPFSTNKGPRILK-----PGEKSTLETKKI 2181  
QY 1391 DNNS--GQSAIKHNKSLITGP-----SKLSRGM--PPELVIPENIKVVERGKYEDYKAGETV 1444  
Db 2182 ESEKGIKGGKVKVYKLLITKTVGRNSEISGOMKQPLQ---ANMPSISRG-----2227  
QY 1445 RSRHSTSVYSSGPSVLRSTLHEAPKALPSGIYDDTSARRTPVSYQNTMGRSGPMNRTSD 1504  
Db 2228 -----RTWIH-----IPGV-----RNSSTSPVSKKGP 2252

QY 1505 VTIPPKNSTHERKSTLTPT--QRESIPAK--SPVPGVDVPSHSPDPHHRGSTAGEVYS 1562  
Db 2253 LKTPASKSPSGQATATSPRGAKFSVKSELSFVARQTSOIGGSSKAFSRSGSR-----2305  
QY 1563 HLPTQLDPAMPFHRALDPAAAYLFQRQLSPTPGYPGSOYQLYAMENTRQTILNDYITSQQ 1622  
Db 2306 -----DSTPSRPAQOPL-----SRP 2320  
QY 1623 MQVNLRPDVA---RGLSPREQPLGLPYPATRGIIIDLNMPTILVPHPGTSTPMDRIT 1679  
Db 2321 IQSPGRNISIPGRNGISPPNK-----LSQLPRT--SPSTASTKSSGSGKMS 2365  
QY 1680 YIPQTQITFPPRPVNSASMSGPHTHLAAAASAEERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRMSQOQLT--KOTGLSKNASSIPRESASKGLNOMNGCANKVELS 2417  
QY 1728 RIAAASDLYLRPGSEGRPGSGHYVRSPPSVRTQETMLQORPSVFOGNGTSTVITPL 1787  
Db 2418 RMSSTKSS-----GSED-----RSERPVLVRQSTFKEAPS-----2449  
QY 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTAADALAALVDAASAPQMDVSKTKESKHEAA 1847  
Db 2450 --PTLR--RKLEESASFESL-----PSSRPASPTRSQAOPTVLSPLDMSLS--THSSVQAGG 2503  
QY 1848 --RLEENLRSSAAVSEOOQLEQKTLVEKRSVQCLYTSSAFFSGKPKQPHSHSVVYSEAGK 1905  
Db 2504 WRKLPNL---SPTEYNDGRPAKRHOIAR-----SHSESPSLPINRSGTWKREHSK 2553  
QY 1906 DKGPPPKSRVEEELRTGKTTITAN-----FIDVITRQIASDKDARERSQSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEDEKHVNSISGTSKOSKNOVSAGKT 2611  
QY 1958 SSSLSHRYE--TPSDAIEVISPASSAPPQEKLTQYQEVVKANQAEENDPTROYEGPLHH 2016  
Db 2612 WRKIKENEFTNTSOTVSSGATNGAESKTLIYOMAPAVSK---TEVWVRIEDCPINN 2668  
QY 2017 YRQOESPSPOQQLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056  
Db 2669 PR---SGRSPGTNTPVIDSVSEKANPNIKDKDQKQKQVNGSVPM--RTVGLENRLN 2723  
QY 2057 QITQDFAHQVSSQTP--QOPTTFQNSPSALVS--TPVTKTTSNRYSPEQASQSVHHQR 2114  
Db 2724 SFIQVADAPQKGTIKPQNNPVPVSETNESSIVERTPFSSSSSKSHSSPSGTVA-----2778  
QY 2115 PGRSVSPENLVDRSGRSGRPGKSPERSHVSSEPEYFISPPQVPPVH--EKQDS 2164  
Db 2779 --ARVTFNY-----NPSPRKSSADTSISARPSQIP--TPVNNNTKKRDS 2818

## RESULT 13

US-08-450-582-7

Sequence 7, Application US/08450582

Patent No. 6114124

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner &amp; Witcoff, Ltd.

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA



ZIP: 20001-4598  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,582  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,655  
FILING DATE: 25-MAY-1995  
APPLICATION NUMBER: US 08/289,548  
FILING DATE: 12-AUG-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/741,940  
FILING DATE: 08-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 1107.49964  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
US-08-450-582-7

Query Match 2.9%; Score 365; DB 3; Length 2843;

Best Local Similarity 19.0%; Pred. No. 6.6e-14;

Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

QY 347 PETKREQOE-RFQVQAGLSATIASEH-----EISEIDGLSEQENNEK 395  
Db 981 PSIESSEDESKFCYGOYPADLAHIAHNSHMDMDGELDTPIYSLKYSDEQLNSGR 1040  
QY 396 QMRQLSVIPPMFDEAORRYKFINMGLMEDPMKYKDRQFMVWTDHE--KEIFDKFI 453  
Db 1041 Q-----SPQNERWARPKHIIIDEIKQSEQRQSRNQSTPYVYTESTDDKHL 1087  
QY 454 QHPKNFGLIASYLERKSVPPCVLYLYLTTRKNENYKALVRNRYGKRRGRNQOIARPSQEEK 513  
Db 1088 KFQPHFG-----QQECVSPY---RSRANGSETNRVGSNHGINQVNSQLQOE- 1132  
QY 514 VEKEEDKAKTEKK--EEKKDEEEDKEDSKENTKEKDKIDGTAEETEERQATPRG 571  
Db 1133 -DDYEDDKPTNYSEYSEEEQHEEERTNYSIKYN-EERHVDQPIDYSLKYATDIPSS 1190  
QY 572 R-----KTANSQGRKGRITRSMTEAAAS-----AAAAATEEPPLPPP 614  
Db 1191 QKSFSKSSSGQSSKTEHSSSESTSTPSSNAKQNLHPSSAOSRSGQPOKAATCK 1250  
QY 615 PEPISTPEVET-----SRTEEMEYAKKGLVEHGRN-----WAAIAKM 653  
Db 1251 VSSINQETIOTCYVEDPICFSRCSLSLSLSSAEDEIGCNOQTQEADSANTLQIAEIKK 1310  
QY 654 VTKSE-----AQCNFYFNKRRINLNLQOHK-----QKTSR 688  
Db 1311 IGRSADBPVSEVPASQHPRTKSSRLQSSLSSESARHRAVEFSSGAKSPSKSGAOTPK 1370  
QY 689 KPREE-----RVSQCESVASTVAQDEDEIEASNEEENPEDSEVAVPSE--DSPEN 740  
Db 1371 SPPEHYVQETPLMFPSRCTSV-SSLDSFESRIASSVOSSEPCSGMVGSIISDLPLDSPGQ 1429

QY 741 A--TSRGNTPEAVELEPTTETAPSTSPSLAVPSTKPAEDSESVETQVINDSISAEAEQMDV 798  
Db 1430 TMPSPRSKTPPP---PPQATKREVPKNAPTAKKESGPKQAANAQVQ-RVQVLPGA 1485  
QY 799 DQOEHSAEE---GSVC-----DPPATKADSDVDVEVR-----VPEN-HASKVECDN 840  
Db 1486 DTLHFATSTPDGFCSSLSALSLSLDEPFIQK-----DVELRIMPPVOENDNGNETESEQ 1541  
QY 841 TKE--RDLDRASEKVEPRDELVVAQINQARPEPQSDNDSSATCSADEVD-----890  
Db 1542 PKESNEQEKAEKTIIDSEKDL-----DDSD-----DDIEIEECIIS 1581  
QY 891 ---GEPERQRMFMDSKPSLLNPTGSLVSSPLKPNPLD---LP-----QLQHRAAVIP 938  
Db 1582 AMPTKSSRKAKKPAQTASKLPPPVAR-----KPSQLPVYKLLPSQNRLQPOKHVSFT 1634  
QY 939 ---PMVSC---TPCNIPIGTPVSGYAL-----YQRHIAKHES 970  
Db 1635 GDDMPRYCVEGTPINFSTATSLTSLDITIESPPNELAAGEVGRGAQSGEFKRTDITPEG 1694  
QY 971 AL-----LEORORQBOIDLECRSSSTSPCGTSKSPNREWEVL-----1007  
Db 1695 RSTDEAOGGKTSVITPELDDNKAEEDILAECLNSAMPKSKHKPRVKKIMDOVOQAS 1754  
QY 1008 --QPAPHOLITNLPPEGVRLPTTRTPPPPLI-----PS 1039  
Db 1755 ASSAPNK---NOLDGKKKKTSPVKPIPONTEYRTRVRKNADSKNNLNAERVFSDNKDS 1811  
QY 1040 SKTYVASEKPSF-----IMGGSIS-----QGTPTGYLISHNOASTYQETPK 1080  
Db 1812 KKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPHHYTIEGTP--YCFSRND-----1860  
QY 1081 PSVGSISLGLPRQOESAKSATLPYIKOEFSRSONSOQPELLVRAOHEGVVRGTAGAIQ 1140  
Db 1861 -SLSLDFDDDDVDLSREKAEKAKENKESAKVTSHTE--LFSNQOSA---NKTQAI 1914  
QY 1141 EGSITRGTPTSKISVESIPSLRGSITQGTPTALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGQPKPIQKQS-----TFQSSKDIIDRGAATDEKLFQNFALIENTPVCFSHNS 1968  
QY 1187 ISRM-----PIEDSSPEKGRREEAASKGHVIEGKSGHI-----1219  
Db 1969 LSSLSLDDOENNNKENEPKIKETEPDQSGEPFKPQASGYAPKSPHFVEDTPVCFGRNSLS 2028  
QY 1220 -LSYDN-----IKNAREGTRSPRTAHEISLKRKSVESVEGNIKQMGMSRESVPVSAPLE 1270  
Db 2029 SLSLSDSEDDLQECISSAMPKKKPSRLKGDNEKSPNNMGILGEDLTLDKDIQRPDS 2088  
QY 1271 GLICRALPRGSPHSDLKERTVLSGIMOGTPRATFESFEDGLKYPKQIKRESPPIRAFEG 1330  
Db 2089 E---HGLSPDSENFDMKAIQEGANSIVSSLHQAAAAAC-----LSRQASDSDSILSILKS 2140  
QY 1331 AITKPKYDGIITIKEMGRSIEHETPRODILTOESRKTPPEVVOSTRPIEGSISOGTPIKF 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPTTSNKGPRILK-----PGEKSTLETAKKI 2181  
QY 1391 DNNS-GQSAIKHNKVLITGP---SKLSRGM-PPLBIVPENIKVBERGKYEDVKAGETV 1444  
Db 2182 ESEKGLKGKGVKYSILITGKVRNSNISGOMKQPLQ---ANMPSISRG-----2227  
QY 1445 RSRHTSVVSGSPSVLRSTLHEAPKAQLSPGIYDTSARRTPVSYQNTMSRGSPMMRTSD 1504  
Db 2228 -----RUMIH-----IPGV-----RNSSSSTSPVSKKGPP 2252  
QY 1505 VTIPPNKSTNHERKSTLTPT-QRESIPAK-SPVPQVDVPSVSHSPFDPHHRGSGTAGEVWS 1562  
Db 2253 LKTPASKSPSEGQATTSRGAKPSVKSELSVARQTSQIGGSSKAPSRSGR-----2305  
QY 1563 HLTPLDPAFPFHRALDPAANAALFQRLSPTPGYPSQYOLYAMENTRQTLINDYITSOQ 1622  
Db 2306 -----DSTPSRAQOPL-----SRP 2320  
QY 1623 MQVNLRPDVA---RGLSPREQPLGLPYPATRGIIDLTMPTILVPHPGGTSTPMDRIT 1679



```
Db 2321 IOSPCRNISPGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGSKMS 2365
Qy 1680 YIPGTQTFPPRPYNASMSGCHTHIAAASAEERERE-----KERERE 1727
Db 2366 Y-----TSPGRMSQOOLN--KOTGLSKNASSIPRSESASKGNOMNNGANKKVELS 2417
Qy 1728 RIAAASDLYLRPGEQGRGSHGYVRSPSVRTOTMLQORPSVFOGNGTGVITPL 1767
Db 2418 RMSSTKS-----GSESD-----RSERPVLRGSTFIKEAPS-----2449
Qy 1788 DPTAQLRIMPLPAGPSISQGLPASRYNTRADALAALVDAASAPQMDVSKTESKHEAA 1847
Db 2450 -PTLR-RKLEESASFESLS---PSSRPASPTRSQAOPTVLSPLDMSLS-THSSVQAGG 2503
Qy 1848 --RLEENLRSAAYVSCQOQLEQKTLVEKRSVOCULTSSAPFSGKQPQHSSVVYSAGK 1905
Db 2504 WRKLPNML---SPTIEYNDGRPAKRHDIAI-----SHSESPSRPLPTRNSGTWKREHSK 2553
Qy 1906 DKGPPPKSRVEEELTRGKTTITAN-----FIDVITRQIASDKDARERGSSQSDS 1957
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEKHEKHVNSISGTKQKENVSAKGT 2611
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPQOEKLTQYQPEVVKANQAEENDPTROYEGPLHH 2016
Db 2612 WRKIKENEFSPTNSTQTVSSGATNGAESKTLIYOMAPAVSK---TEDVMVRIEDCPINN 2668
Qy 2017 YRPOEESPSPOQILPP-----SSQAEQMGQVPRTHRLITLADHIC 2056
Db 2669 PR---SGRSPTGNTFPVIVDSYSEKANPNIKSDKNQAKQNGVSPM--RTVGLNRLN 2723
Qy 2057 QIITQDFARNQVSSQTP--QOQPTSTFQNSPVALYS--TPVTRKTSNRYSPESQAOVSHQR 2114
Db 2724 SFIQVDAPDQKTEIKQCONNPVPVSETNESSIIVERTFFSSSSSKHSSPSGTVA-----2778
Qy 2115 PGRSVENLVKSGRSGRPSKPSERSHVSSEPYEPIPPQVPPVHH--EKQDS 2164
Db 2779 --ARVTFPNY-----NPSRKSSADSTSARPSQIP-TPVNNNTKKRDS 2818
```

## RESULT 14

US-08-821-355A-7  
Sequence 7, Application US/08821355A  
Patent No. 5851775

## GENERAL INFORMATION:

APPLICANT: Barker, Nick  
APPLICANT: Clevers, Hans  
APPLICANT: Korinek, Vladimir  
APPLICANT: Morin, Patrice  
APPLICANT: Kinzler, Kenneth  
APPLICANT: Vogelstein, Bert  
APPLICANT: Sparks, Andrew  
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
TITLE OF INVENTION: Interact to Prevent Cancer

## NUMBER OF SEQUENCES: 11

## CORRESPONDENCE ADDRESSES:

ADDRESS: Banner & Witcoff, Ltd.

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20001

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/821,355A

FILING DATE: 20-MAR-1997

CLASSIFICATION: 514

## PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32,145  
REFERENCE/DOCKET NUMBER: 1107.05064  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 97430 BMB UT  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2973 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5851775e  
US-08-821-355A-7

Query Match 2.9%; Score 365; DB 2; Length 2973;  
Best Local Similarity 19.0%; Pred. No. 7.1e-14;  
Matches 409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;

```
Qy 347 PEIKKOREQOE-RFORVGQRGAGLSATIRSEH-----EISEIIDLGLSEQENNEK 395
Db 981 PSIESYSEDDSKFCSYGYQYPADLAHKHSANHMDNDGELDTPIYSLKYSDEQLNSGR 1040
Qy 396 QMRQLSVIPPMFPAEORRYKFINNGLMEDPMKYKDRQFMNVWTDHE--KEIFKDKFI 453
Db 1041 Q-----SPSQNERWARPKHIEIETIKQSQRSRQSTTYPVYTESTDDKHL 1087
Qy 454 QHPKNFGLIASYLERKSVDPCLVYLYLTKNENYKALVRRNYGKRRGRNQOIARPSOEK 513
Db 1088 KQPHFG-----QQECVSPY----RSRGANGSETNRVGSNHNQINQVNSLQOE- 1132
Qy 514 VEEKEEDAKAEKTEKK--EEKKDEEKEDEKEDSKENTKDKIDGTAEETEEREQATPRG 571
Db 1133 -DYEDDKPTNYSERYSEEEQHEEERTNYSIKYN-EKRRHVDPIDYSLKYATDIPSS 1190
Qy 572 R-----KTANSQGRKGRITRSMNTNEAAS-----AAAAAATEEPPPLPPP 614
Db 1191 QKOSFSGKSSGSSKTEHMSSESTSTPSSNAKRONLHPSSAQSRSQGPQKAATCK 1250
Qy 615 PEPITSEPVET-----SRWTEEMEYAKKGLVHGHN-----WAAIAKM 653
Db 1251 VSSINQETIQTICVEDTPICFSCSSLSLSAEDGICNQTQEQEADSNANTLQIAEIKK 1310
Qy 654 VGTKSE-----AQCKNFYFNKRRHNDNLQOQK-----QKTSR 688
Db 1311 IGTRSAEDPVSEVPASVQHPRTKSSRLQSGSLSESARHRAVEFSGAKSPSKGAQTPK 1370
Qy 689 KPREE-----RVSOCESVASTVSAQEDEDIASNEENPEDESEVAVKPE--DSPEN 740
Db 1371 SPPEHYVQETPLMFSTRCTSV-SSLDSFESIASVSSQSEPCSGMVSGIISPSDLPSPGQ 1429
Qy 741 A--TSRGNTPEFAVELEPTTETAPSTPSLAVPSTKPAEDESVEFQVNDSDISAETAQMDV 798
Db 1430 TMPSPRSKTPPP---PPQATQREVKNKAPTAKEKESGPKQAANVAQV-RVQVLPDA 1485
Qy 799 DQOEHSAAE-----GSVC-----DPPPATKADSVDEVR-----YPEN-HASKVGEDN 840
Db 1486 DTLHLHFATESTPDGFCSSLSLSALSLDEPFIQK---DVELRIMPPVQENDNGNETESEQ 1541
Qy 841 TKE--RDLDRASEKVEPRDEDLVVAQOINAQRPPEQSDNDSSCATSADQVD-----890
Db 1542 PKESNENQEKAEKTIIDSEKDLL-----DDSD-----DDDIETILEECIIS 1591
Qy 891 ---GEPERQRMFMDSKPSLLNPTGTSILVSSPLKPNPLD---LP-----QLQRAAVIP 938
Db 1582 AMTKSSRKAKKPAQATASKLPPPVAR-----KPSOLPVYKLLPSQNRLOPKQHVSTFP 1634
Qy 939 ----PMVSC---TPCNIPITGTPVSGYAL-----YORIKAHMES 970
```



Db 1635 GDMPRVYCVETGPIINFSTATSLDLTIESPPNELAAGEVRGGAQGEFEKRDITPTIG 1694  
Qy 971 AL-----LEORQOEIDLECRSSSTPCGTSKSPNREWEVL----- 1007  
Db 1695 RSTDEAOGGKTSSVTIPELDNKAEEGDILAECINSAMPKSHKPRVKKIMDOVOQAS 1754  
Qy 1008 --QPAPHOLITNLPPEGVRLPTRRPPPLI-----PS 1039  
Db 1755 ASSAPNKK--NQLDGKKKKTPSPVKPIQONTVEYRVRKNADSKNLLNAERFVSDNKDS 1811  
Qy 1040 SKTTVAESEKPSF-----IMGGSIS-----OQTPGYLTHSHNOASTQETPK 1080  
Db 1812 KQQLKNNSKDFNDKLPNNEDRVGRFAFDSPHHYTPIEGT--YCFSRND----- 1860  
Qy 1081 PSVGSISLGLPQOESAKSATLPYIKOEFSPRSONSOPEGLLVRAOHEGVVVRGTAGAIQ 1140  
Db 1861 -LSLSLDFDDDDVLSREKAELEKAKENKESEAKVTSHT--LTSNQOSA---NKTQAI 1914  
Qy 1141 EGSTTRGTPTS KISVESIPSLRGSIOTGPALPOTGIPTEALVK-----GS 1186  
Db 1915 KQPINRGOPKPILOKQ-----TFQSSKDIPDRGAATDEKLQNFALIENTPVCFSHNS 1968  
Qy 1187 ISRM-----PIEDSSPEKGREREAASKHVIYEKSGSHI----- 1219  
Db 1969 LSSLSIDDOENNKNENEFIKETEPDQSGEPKQASGYAPKSFHVEDTPVCFSRNSSL 2028  
Qy 1220 -LSYDN-----IKNAREGTRPRTAHEISLKRVSYESVGNIKOGMSRESVPSAPLE 1270  
Db 2029 SLSIDSEDDLLQECISSAMPKKPSRLKGDNEKHSRPNMGIIIGEDLTDLKDIPRDS 2088  
Qy 1271 GLICRALPRGSHDLKERTVLGSIOMGTTPRATESFEDGLKYPKQIKRESPIRAFEG 1330  
Db 2089 E---HGLSPDENFDWKAIOEGANSIVSSLHQAAAAAC-----LSRQASSDSLSLSKS 2140  
Qy 1331 AITGKPYDGTITIKEMGRSHEITPRQDILTQESRKPEVVOSTRPIEGSISOGTPIKF 1390  
Db 2141 GISLGSFP-----HLTPDQEEKPFTSNKGPRIKL-----PGEKSTLETKKI 2181  
Qy 1391 DNNS-GOSAIAHNKVSILITGP-----SKLSRGM-PPLEIVPENIKVWERYEDVKAGETV 1444  
Db 2182 ESESGIKGKKVYKSLITKVRNSSEISGOMKQPLQ---ANMPSIRG----- 2227  
Qy 1445 KSRHTSVVSGPSVRLSTLHEAPKAQLSPGIYDDTSAARTPVSYQNTMSRGSPPMNRSTD 1504  
Db 2228 -----RTMIH-----IPGV-----RNSSSSTSPVSKGKGP 2252  
Qy 1505 VTIPPKNSTNHERKSLTPT-QRESIPAK-SPVPGVDPVYVSHSPFDPHHRGSTAGEVYWS 1562  
Db 2253 LKTPAKSPSEGOATTSPRGAKPSVKSELSPVARQTSQIGGSKAPSRSGR----- 2305  
Qy 1563 HLPQOLDPAMPFHRALDPAAYLFQRLSPTPGYPSQOYLYAMENTROTILNDYITSQ 1622  
Db 2306 -----DSTPSRPAQPL-----SRP 2320  
Qy 1623 MQVNLRPDVA---RGLSPREOPLGLPYATRGIIIDLNMPTILVHPGGSTPPMDRIT 1679  
Db 2321 IQSPGRNISPGRNGISPPNK-----LSQLPRT-SSPSTASTKSSGSGKMS 2365  
Qy 1680 YIPGTQITFFPRPNYSASMSGPHTHLAAASAERERERE-----KERERE 1727  
Db 2366 Y-----TSPGRQMSQNLIT--KOTGLSKNASSIPRSESASGLNQMNNGANKRVLS 2417  
Qy 1728 RIAAASDLYLRPGSEQPGRPGSHGVYRSVSPSVRTQETMLQQRPSVFGQNGTSVITPL 1787  
Db 2418 RMSSTKSS-----GSESD-----RSERPVLVROSTFIKEAPS----- 2449  
Qy 1788 DPTAQLRIMPLPAGGPIISQGLPASRYNTAADALAALVDAASAPQMDVSKTESKHEAA 1847  
Db 2450 -PTLR-RKLEESASFELS---PSSRPASPTRSQAQTPVLSPLPDMSL--THSSVQAGG 2503  
Qy 1848 --RLEENLRSAVSEQQOLEQKTELEVERKSVOCLTSSAFPSGKQPQSHSVVYSAGK 1905  
Db 2504 WRKLPPNL---SPTIYNDGRPAKRHDIAI-----SHSESPSRLPINRSKGTWKREHSK 2553

Qy 1906 DKGPPKSRYEELRTRGKTTITAN-----FIDVILITRQIASDKDAREGSSSDS 1957  
Db 2554 HSSSLP--RVSTWRTGSSSILSASSSESEKSEKSEKHYNSISGTSQKRENOVSAGKT 2611  
Qy 1958 SSSLSHRYE-TPSDAIEVISPASPAPQBEKLOTYQEVVYKANAENDPTRQYEGPLHH 2016  
Db 2612 WRKIKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668  
Qy 2017 YRPOQESPSPOQLPP-----SSQAEQMGQVPRTHRLITLITLADHIC 2056  
Db 2669 PR-----SGRSPTGNTPPVIDSVSEKANPNIKDSKONQAKQNGSVPM--RTVGLENKLN 2723  
Qy 2057 QIITQDFARNQVSSQTP--QOPPTSTFQNSPVALYS-TPVRTKTSNRYSPESQAOQSVHHQR 2114  
Db 2724 SFIQVDAPDQKGTETKPCQNNPVVSETNNESSIIVERTPFSSSSSKHSSPSGCTVA----- 2778  
Qy 2115 PGRSVSPENLVKSRGSRGKSPERSHVSSEPIFISPPQVPPVH--EKQDS 2164  
Db 2779 --ARVTPFNY-----NPSPRKSSADSTSARPSQIP-TPVNNNTKKRDS 2818

## RESULT 15

US-09-003-687A-7  
; Sequence 7, Application US/09003687A  
; Patent No. 5998600  
; GENERAL INFORMATION:  
; APPLICANT: Barker, Nick  
; APPLICANT: Clevers, Hans  
; APPLICANT: Korinek, Vladimir  
; APPLICANT: Morin, Patrice  
; APPLICANT: Kinzler, Kenneth  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Sparks, Andrew  
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 1001 G Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/003,687A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/821,355  
; FILING DATE: 20-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32,145  
; REFERENCE/DOCKET NUMBER: 1107.05064  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX: 97430 BMB UT  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2973 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5998600e  
US-09-003-687A-7



	Query Match	2.98; Score 365; DB 2; Length 2973;
	Best Local Similarity	19.0%; Pred. No. 7.le-14;
	Matches	409; Conservative 288; Mismatches 807; Indels 648; Gaps 94;
QY	347 PEIRKOREQE-RQRVQCORGAGLSATARSEH-----EELSEIIDGLSEGENNEK	395
DB	981 PSIESYEDDESKSCSYGOYPADLAHKTHSANHHMDNDGELDTPINYSLKYSDLEOLNSGR	1040
QY	396 QMRQLSVPPMMFAEQRRVFNNNGLIMPMVKYDKRQFMNVWTDHE--KEIFPKDFXI	453
DB	1041 Q-----SPSQNERWARPKHIIEDEIKOSEQORSRNQSTTVPVTESIDDKHL	1087
QY	454 QHPKNFGLIASYLERSKSVDPDCVLYYLYTKNENYKALVRNYGKRGRNQIARPSQBEK	513
DB	1088 KFQPHFG-----QQECVSPY-----RSRGANGSETNRVGSNHGNQVNSQLCQE-	1132
QY	514 VEEKEEKAEKTEKK--REEKKDEEEKDEKEDSKENTKEKDIDGTAAETEEBEOATPRG	571
DB	1133 -DDYEDDKPTWYSRYSEEQHEERETPNYSIKYN-EEKRHVQPIDISLUKYATDIPSS	1190
QY	572 R-----KTANSQGRRRKGRITRSMTNEAAAAA-----AAAAAABEPPLPPP	614
DB	1191 KQQSFSFKSSGGOSKTEHSSSENTSTPSSNAKRQNQLHPSSAQSGOPQAATCK	1250
QY	615 PEPITSEPVE-----SRTEEEMEVAKGLVEHRN-----WAAIAKM	653
DB	1251 VSSINQETIQTYCEDTPICFSRCSSLSLSAEDAIGCNQTTOADSANTLIQTAEIKEK	1310
QY	654 VGTKSE-----AOCKNFYNYKRRNLNLLQOHK-----QKTSR	688
DB	1311 IGTSAGEDPVSEVAVSQHPTKSRLOGSSLSSASARKHAVESPSSGAKSPKSQAQTPK	1370
QY	689 KPREE-----RDVSOCSVASITYSAODEDIEASNEEENPEDSEVAVKPSE--DSPEN	740
DB	1371 SPPEHYVOETPLMFSRCTSV-SLSDSPESRSIASSVQSEPCSGMVGSIISPDLPSDQG	1429
QY	741 A--TSRGNTPEAVELEPTETAPSTPSLAVPSTKPADESVEFYQVNDISIAETAQMDV	798
DB	1430 TMPSPSRSKTPPP---PPQTAOTKREVPKNKATAEKRESGPQKAANAQV-RQVLPDA	1485
QY	799 DQOHSABE-----GSVC-----DPPPATKADVDEVR---VPEN-HASKVEGDN	840
DB	1486 DTLLHFATESTPDGFSCSSLSALSDELFFQK---DVLRIMPVQENDNGNETESEQ	1541
QY	841 TKE--RLDLDRASEKVPREDLVVAQQINAQRPBPQSDNDSATCSADEDVD-----	890
DB	1542 PKESNENOKEAKETIDSEKDLL-----DDSD-----DDDTEILLEECIIS	1581
QY	891 ---CEPERQRMFPMDSKPSLLNPTGSIILVSPLKPPLD---LP-----OLOHRAAVIP	938
DB	1582 AMPTKSRKAKKPAQATASKLPPVAR-----KPSQLPVYKLPLPSQNRLQPQKHVSFT	1634
QY	939 ----PMWSC----TPCNTPIGTGPVSGYAL-----YORHIKAMHES	970
DB	1635 GDDMPRYCVEGTFINSTATSLSDLATIESPNELAAGEGVGGAQSGEFKRTIPTEG	1694
QY	971 AL-----LEBRQRQEQJIDUECRSSTSFCGTSKSPNREWYL-----	1007
DB	1695 RSTDAGCGKTSSTVPIPELDNDKNAEGLAEACINSAMPKGKSHKPFVKKIMQVQOAS	1754
QY	1008 --QPAPHOLITNLPEGVRLPTRTRTPPPPLI-----PS	1039
DB	1755 ASSGAPNK---NQLDGKKKTKSPVKPIQNPQTEYTRVRKNADSKNNLNAREVPSDNKDS	1811
QY	1040 SKTTVASEKPSF-----IMGGIS-----QGTPTGYLTLSHNOASYTQETPK	1080
DB	1812 KKONLKNNSKDFNDKLPNNEDVRVGSFAFDSPHHVPTIEGTF--YCFSRND-----	1860
QY	1081 PSVGSISLGLPRQBSAKSATLPYIKOEFPSPQRSNQSOGPELLVRAHQEVGVRTAGAIQ	1140
DB	1861 -SLSSLDQDDDDVLSREKABLREKAKENKESAKVTSHTS--LTSNQOSA---NKTOAIA	1914







```
Qy 1728 RIAAASDLYLRLPSEQPGRGSHGYVRSPSPSVRTQETMLQQRPSVFGQTNGTSVITPL 1787
Db 2418 RMSSTKSS-----GSESD-----RSERPVLRQSTFIKEAPS----- 2449
Qy 1788 DPTAQLRIMPLPAGGPSISQGLPASRYNTADALAAALVDAASAPQMDVSKTKESKHEAA 1847
Db 2450 -PTLR-RKLEESASFESLS---PSRPASPTRSOAQTPLVSLPDMASLS-THSSVOAGG 2503
Qy 1848 --RLEENLRSAVSEOOEQLEQTEVEKRSVOCCLYTSSAFPSGCKPQPHSHVVYSEAGK 1905
Db 2504 WRKLPPNL---SPRIEYNDGRPARHDIAR-----SHSESPRLPNRSGTWKREHSK 2553
Qy 1906 DKGPPPKSRYEEELRTKTTITAA-----FIDVIITQIASDKDAREGQSQSDS 1957
Db 2554 HSSSLP--RVSTWRTGSSSILSASSESEKAKSEDEKHVNSISGCKQSKENQVSAKGT 2611
Qy 1958 SSSLSHRYE--TPSDAIEVISPASPPQBEKLTQYQEVVYKANOANDPTROYEGPLHH 2016
Db 2612 WRKIKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSK---TEDVWVRIEDCPINN 2668
Qy 2017 YRPOQSPSPQOOLPP-----SSQAEQMGQVPRTHRLITLADHIC 2056
Db 2669 PR-----SGRSPGTNTFPVIDSVSEKANPNIKDKNQAKQNGSVPM--RTVGLNRLN 2723
Qy 2057 QIITQDFARNQVSSQTP--QOPPTSTFQNSPALSYS--TPVRTKTSNRYSPESQAQSVHQR 2114
Db 2724 SFIQVDAPDQKTEKPCQNNPVPVSETNESSIIVERPFSSSSSKHSSPSGTVA----- 2778
Qy 2115 PGRSVSPENLVKSRGSRGPRGSPERSHVSSEPYEPISPQVPPVHH--EKQDS 2164
Db 2779 --ARVTPFNY-----NPSRKSSADTSARPSQIP-TPVNNNTKKRDS 2818

RESULT 17
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Sellerli, Licla
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
```

```
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-061-376-5

Query Match 2.6%; Score 329; DB 4; Length 3969;
Best Local Similarity 17.5%; Pred. No. 1.8e-11;
Matches 593; Conservative 378; Mismatches 1089; Indels 1332; Gaps 153;

Qy 1 MSSSGYPPNQGAFTGOSRPPPHSVQYTFPNTRRHQQEFV-----PDYRSHL 48
Db 278 LSPSSLSLROGSEFKZGEGYKLTZTERKASINRKDKDPFGLLIILNWKSPRKSQKTRKHL 337
Qy 49 E-----YSOASQLLQOQOQOOLRRRPSLLSEFHPGSDRPOBERTSYEPFH 93
Db 338 HLQKKIRQLSDKALEGLSOLGLFLQKGQMO-----PLLSNSYRGQKGAQKKEKA-- 390
Qy 94 PGSPVDHDSLESKRPRLEQVSDSHFORVSAAVLPLVHPLPEGL-----RASADAKDP- 147
Db 391 -----AQLQGRKVK-----TQVKNIQOFIMPVVSIAISSRIIKTPRRFIEDYDPP 436
Qy 148 -AFGKKHARSSPISGQPCGDDQNAS-----PSKLSKEELIQMDRVDREI---AKVEQQ 198
Db 437 IKIARLESTPNRFSAPS CGSSEKSAASQHSQSSSSPSVDITSTDSQASEIQ 496
Qy 199 IL-----KLVKKQOQLEEEA 213
Db 497 VLPERSDTPVHPPLPISQSPNESNDRRRYSVRSERSFGSRTTKLSTLQSAPOQOT 556
Qy 214 AK-----PPEPEKPV-----PP-----PVEQKHSIV----- 236
Db 557 SSSPPPLLTTPPLQPASSISDHTPWLMPPTIPLASFPPLPASTAPMGKRSILREPTF 616
Qy 237 -----QIYDENRKAEEA---HKIFEGLPKVELPLYNQPSDT----- 272
Db 617 RWTSLKHSRSEPQYFSSAKYAKEGLIRKPIFDNFRPP---PL--TPEDVGFASGSASGT 671
Qy 273 -----KVYHENIKTNQVMRKKLI---LFFKRNHARKQREKICORYDOLMEAWKKV 322
Db 672 AASARLSPHSGTRFDMHKRSPLLRAPRTPPSEASR-----IFESVTLPS 718
Qy 323 DRINNPRKAKESKTREYKEKQPEIRKQEQOERFORVQORGAGLSATIAREHEISE 382
Db 719 NRTSAGTSSSGSVNRKRR--RKVESPIRSPSPSHSMRT--RSGRLSSSELSPLTPSS 774
Qy 383 IIDGLS-----EQENNEKQMRQOLSVPPPMWFAEQRRV 415
Db 775 VSSLSISVSPSATSNLPTFTFPFSLTSGESAENQRPRKQTSAPAEFFSSSSPTP 833
Qy 416 KF-----INMGLMED--PMKVYKDRQFMVMTDHEKEIFKDKF----- 452
Db 834 LFPWFPTGQTERGRNKKDKAPELSKDR-----DADKSVEKDKSRDREREKENRE 886
Qy 453 -----IQHPKNFGLIASYLERKSVDPDCVLYYLTKNE----- 485
Db 887 SRKEKRRKGSIEQSSSALYPVGRVSKYKGVGEDVATSSAKKATGRKKSSHSDSGTDITS 946
Qy 486 -----NYKALVRRNYKRRGRNQOILARPQOEKVEEK----- 517
Db 947 VTLDGTTAVTKILIKKGRGNLEKTNLDLG-PTAPSLKEKTKLCLTSPSSSTVHKHSTSI 1005
Qy 518 -----EEDKAEKTEKKE-----EEKDEEEKDEKEDSKENTKEKDKIDGTABETEEREQAT 568
Db 1006 GSMLAQADKLPMWTDKRVASLLKKAQICKIEKSKLQOT-DQPKAQQESDSSETSVRG 1064
Qy 569 PR-----GRKTANSQGRKG-----RITPSWTNEAAASAAAAATEE 606
Db 1065 PRIKHVCRRAAVALGRKRAVFPDMDPTLSALPWPEREKILSMGNDKSSTAGS-----E 1119
Qy 607 PPPPLPPPPPISTEPVETSRWTEEMEVAKKGLVHGHNWAAIAKMWGTSKSAQC----- 662
Db 1120 DAEPLAPPKPII--KPVTRNKAPQEP-----VKGRSRRCGCGPGCVPEDCGCT 1170
```







Db 3089 -----VLQTLPLNGVTKIQLTSSVSTPSVMTNTSV---LGPWGGGLTLTLTGLNP--- 3136  
QY 2235 IFNLPAVTTSGVSRG-----HSFADPASNLGLEDIIIRKALMGSDFDKVEDHGV 2284  
Db 3137 --SUPTSLSLPSASKGLLPSHHQHLHSF--PA-----ATQSSPPPNIS----- 3177  
QY 2285 VMSOPMGVPGT---ANTSVTSGETREED---PSPHSGGVCKPKLISKSNSRKSXP 2338  
Db 3178 --NPPSGLLIGVQPPDPOLLVSESSQRTDSTVATPSSG--LKKRPISRLQTRKKKL 3233  
QY 2339 IPGQYLGTERPSV--SSVHSEGDYHQTPGMAWEDRPSSTGSTQPPNPLTMRMLSS 2396  
Db 3234 AP-----SSTPSNAPSDDVSNMTLNTF-----SQLPNHP-SLLDLGSL 3273  
QY 2397 PPTIACAPSAVNAAPHQNRWEREPAPLL 2428  
Db 3274 NTSSHRTVPNIKRS-----KSSIMYFEPAPLL 3301

RESULT 18  
US-08-769-309A-5  
; Sequence 5, Application US/08769309A  
; Patent No. 5741890  
; GENERAL INFORMATION:  
; APPLICANT: Scott, John D.,  
; APPLICANT: Nauert, Brian J.,  
; APPLICANT: Klauck, Theresa M.,  
; TITLE OF INVENTION: Protein Binding Domains of Gravin  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/769,309A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5741890and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/33451  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1780 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-769-309A-5

Query Match 2.6%; Score 323.5; DB 1; Length 1780;  
Best Local Similarity 18.6%; Pred. No. 1.3e-11;  
Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

QY 496 KKRGRNQIARPSQEKV-----BEKEDKAEKTEKKEEKKDEEKDEKDSKENT 548  
Db 83 GQKALNGOALNSQEEIEVTEVGQDSSEVSDSKEMATKSAVVDITDQGEEN 142  
QY 549 KEKDIDGTAEETEEREQANPRKRTANSOGRRK-----GRITR 587  
Db 143 RNIBQIPSSSENLEELTQPT---ESQANDIGFRKRVFVFGFKFTVKDKTEKPDVTQVLLT 199

QY 588 SMTNEAANAASAA-----AAAATEPPPLPPPPPISTE-----PVETSR 627  
Db 200 VKDEGEAGAGDHQDPSLGAEEAAKSEKPEKOSTEKPEETLKREOSHABISPAESGQ 259  
QY 628 WTEF-EMEAVAKGLVHGRNMAAIKAVGVTKSEACQCNFYFNKRRHNLNDLLOHQHKQT 686  
Db 260 AVECKEKEGEKQEKESKSAESPTSPVTSSTGTFKFF-----TQWAGWRKTS 311  
QY 687 SRKPREDRVSQCESVASTVSAODEDIEASNEENPEDSEVEAVKPSDEPENATSRGN 746  
Db 312 FRKPK-----DEVEASEKKKEQEKVDTEEDGAEVASEKLTASEOAHQPOEPAESA 364  
QY 747 TEP-----AVELEPTTETAPSTPSLAVPSTPAE--DESVE----- 781  
Db 365 HEPRLSAEYKVELPSBEQVSGSQGPSEKPAPLATEVFEDEKIEVHQEEVVAEVHVSTVE 424  
QY 782 -----TQVND---SISAETAEMDVDOEH-----SAEESVCDPPPPATRAD--SVD 823  
Db 425 ERTEEQKTEVEETAGSVPAEELVCMDAEPQEAEPKELVKLKEKTCVSGEDTQCADLSPD 484  
QY 824 VEV--RVPENHASKVEGDNTKER-----DLDRASEKVEP-----RDEDLV 861  
Db 485 EKLSKPPGEGVYSEVEMLSQOERMKVQGSPLKLFSTGLKLLSGKKQKGRGGDEESG 544  
QY 862 VAQOINARPEQSDNDSSATCSADE-----VDEPERQ 896  
Db 545 EHTQVPADSPDSEEQKGESSASSPEPEITCLEKGLAEVQOGEAEBGATSGEKKRE 604  
QY 897 RMFPMDSKPLLNPSTGILVSSPLKPNPLDLQLOHRAAVIPPMVSVCTPCNIGTIPVSG 956  
Db 605 GVTPWASFKKWTP----- 618  
QY 957 YALYQRIHKAHESALLEBQROEQIDLECRSTSPCTGSKSPNREWEVLQAPHQLIT 1016  
Db 619 ----KRVRRPSES-----DKEDLDKVKSATLSSTESTASEMOEEMKSGVEEPAPEE--- 667  
QY 1017 NLPEGVRLPTRTRPPPLIPSKTTVASEKPSFIMGSSISQCTPGTYLTSHNQASYQ 1076  
Db 668 -----PRKVDTSVSEALICVSGSKK-----RARRSSSDE 699  
QY 1077 ETPKPSVGSISLGLPROQESAKSATLYIKOEFPSPQNSQPEGLVLVRAQHEGVVRGTA 1136  
Db 700 EGGPKAMGG-----DHQRADE-AGDKKETGTDGILAGSOEHDPGGQSS 741  
QY 1137 GATQEGSITRGTPTSKISVESIPSLRGSITQTPALPOTGIPTALYKGSISRMPEDSS 1196  
Db 742 SPEQAGS-----PTEGEGVSTWESFKRLVTPRKSKSLEEKSEDSIAGS-----GVEHST 792  
QY 1197 P--EKGREEAASKGHVIEGKSGHLSYDNIKNAREGTRSPRTAHEISLKRYSYSEVGN 1254  
Db 793 PDTEPGKEE-----SWVSIIKFIPIRRKKRP---DGKQEQAPVEDAG 831  
QY 1255 KQMSMRSPVSA--PL-----EGLICRALPRGSPHSDLKERTVLVSGSIMOGTPTAT 1304  
Db 832 PTGANEDSDVPVAVVPLSEYDAVEREKMEAQAQKAEQEAQAAATEVSELSQSQVHMM 891  
QY 1305 TESFEDGLKYPKQIKRESP-----PIRAF--EGAITKGRPYDGTITIKEMGRSIEH 1353  
Db 892 AAADVADGTRAATIEERSPSWISASVTEPLEQVPAEALITEEVLEVEVIAEEEPPTVE 951  
QY 1354 -IP-----RODILTQESRKTPVYVQSTRPIEGSISQCTPIKFDNNSGQSAIKHN----- 1402  
Db 952 PLPENREARGDVTVSEAEITPEAVTAA-----ETAGPLGSEEGTEASAAEETEMVS 1003  
QY 1403 -VKSITGPKSLSRGMPLLEI---VPENIKVVERGKYEDVKA-GETVRSRHTSVVSSGP- 1456  
Db 1004 AVSOLTDSPDPTTEATPVQEVGVPD-IEEQERRTOEVLQVAEAKVKEESQPLGTPGPE 1062  
QY 1457 SVLRSTLH---EAPKAOL-SPGIYDDT-----SARRTPVSYQNTMRSGSP-MMRT 1502  
Db 1063 DVLQPVORAEARPEQEAESGLKKETDVLVVDQAQEAKEPTTQGVVGGTTPTESFEKA 1122



Qy 1503 SDVT--IPNKNSTHNRKSTLTPTQRESIPAKSPVPGVDPVSHSPFPHRGSTAGEVY 1560  
Db 1123 PQVTIESSELVTTQCAETLAVGKQBMWQAIP--PDSVETPTDSETDGTTPVADF 1179  
Qy 1561 WSHLPTQLDPAMPFHRA---LDPA---AAAYLFQROLSTPTGYPQYQIYAMENTROT 1612  
Db 1180 DAPGTTQKDEIVHEENHVLVVRGTEAEVPAQKERPPA---PSSF-VFQETKREQS 1235  
Qy 1613 ILNDYITTSQOQVNLRPDVARGLSPR-----EPLGLP-YPATRGIDLTNNPP 1660  
Db 1236 KMEDTLEHTKEVSV--ETVSILSKTEGTQADQADYEKTKDVFFEGLEGSID-TGI-- 1290  
Qy 1661 TILVPHPGGTSTPPMDRTYI---PGQIITFPPR---PYNASMSPHGPHTHLAAAASA 1712  
Db 1291 -----TVSRKTEVVALGEGTEAECKDDALELQSHAKSPSPVEREMVQV 1339  
Qy 1713 ERER-ERE---REKERERERIAASDLYLRPGSEQ-----PGRGSHGYVRSP 1757  
Db 1340 EREKTEAEPTHVNEEKLHETAVTVSEEV-----SKQLLOTVNVPIIDGAKEVSSLEGSP 1394  
Qy 1758 SPVVRTQETM-----LQORPSVFGCTNGTSVITPLDPTAQLRIMPL-----PAGGPSISQ 1807  
Db 1395 PPLCQOEAVCTKIQVQSSSEASFILTAAREEKEVLGETA--NILETGETLEPAGAHVLVE 1452  
Qy 1808 GLPASRYNTAA-----DALAALVDAAASAPQMDYS-----KTKESKHEAAR 1848  
Db 1453 EKSEKNEEDFAHPGEDAVPTGPDCAKSTPIVYSATTKKGLSSDLEGEKTTSLKWSDE 1512  
Qy 1849 LEENLRSRAVVS---EQOQLEQKYLEVKKRS-----VOCLYTSSAPPSG 1890  
Db 1513 VDEQVACQEVKSVVAIEDLEPENGILELETKSKLVQNIQTAVDQFVTRTEATEMLTS 1572  
Qy 1891 KPQHSSVYSEACKDKGPPKPSRYEE--ELTRGKTTITTAANFDIVITTRQIASDKDARE 1949  
Db 1573 ELQTAHVIKADS--QDAQETKEGEEEPQASAQDETPIISA----- 1612  
Qy 1950 RGSQSSSSSLSHRYETPSDAIEVISPASSPAPPQKLOTQYQVVKANQANDPTRQ 2009  
Db 1613 --KEESTAVGAQH-----SDISKMSSEAS-----EKTMTVEVE-----GSTVND--QQ 1653  
Qy 2010 YEGPLHYRPOQSPSPQOQLPPSSQABGCMG--QVPRTHRILITLADHICQIITO----- 2061  
Db 1654 LE-----EVLWPSSEEGGAGTKSVPEDDGHALLAERIEKSLVPEPKEDEK 1698  
Qy 2062 ----DFARNQVSS-----QTPQOPPTSTFONSPS-----ALVSTPVTKTS 2098  
Db 1699 GDDVDDPENQNSALADTASGLTKESPD-----NGPKQKEKEDAQEVLEQEGKVHSESD 1754  
Qy 2099 NRYSPESQAQSVHHQRFGSR 2118  
Db 1755 KAITPQAQEELOKQERESAK 1774

## RESULT 19

US-08-994-570-5

Sequence 5, Application US/08994570

Patent No. 6090929

GENERAL INFORMATION:

APPLICANT: Scott, John D.,

APPLICANT: Nauert, Brian J.,

APPLICANT: Klauck, Theresa M.

TITLE OF INVENTION: Protein Binding Domains of Gravin

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Marshall, O'Toole, Gerstein, Murray &amp; Borun

STREET: 6300 Sears Tower/233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: United States of America

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/994,570

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: No. 6090929and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/33451

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-994-570-5

Query Match

Best Local Similarity 2.6%; Score 323.5; DB 3; Length 1780;

Matches 361; Conservative 273; Mismatches 741; Indels 565; Gaps 82;

Qy 496 KGRGRNQIARPQEEKV-----EKEEDKAKTEKKEEKKDEEKEDEKESKENT 548  
Db 83 GQKALNGQGALNSQEEVIVTEVQGRDSESDVSEKSDKEMATKSAVVHIDTDGQEN 142  
Qy 549 KEKDKIDGTAEETEREQATPRKRTANSQGRK-----GRITR 587  
Db 143 RNIIQIPSESNLELTQPT---ESQANDIGKVKFKVGVKFTVKDKTKTKPTVQLLT 199  
Qy 588 SMTNEAAAAA---AAAAEPPPPPLPPPPPISTE-----PVETSR 627  
Db 200 VKDEGEAGAGDHDPSLAGEAASKESEPKQSTKEPTLKRQSHAEISPPAESGQ 259  
Qy 628 WTER-EMEVAKKGLVEHGRNWAATAKVVGTSEACKNFYFNYKRRHNLNLLQHKOKT 686  
Db 260 AVECKEKEGEEKPEKSKAESPTSPVTSSTGTFKFF-----TOGWAGWRKKTS 311  
Qy 687 SRKPREEDVQCESVASTVSAQEDIEDIASENEENPDSEVAVKPSDESPENATSRN 746  
Db 312 FRKPE-----DEVEASEKKKQEPKVDTEEDGKAEVASEKLTASEQAHQPPAES 364  
Qy 747 TEP-----AVELEPTTETAPSTSPSLAVPSTKPAE--DESVE----- 781  
Db 365 HEPRLSAEYKVELPSEFQVSSQGPSEKPAPLATEVFDEKIEVHQEVVAEVHVSVE 424  
Qy 782 -----TQVND---SISAEAEQMDVDQEH-----SABEGSVCDPPPATKAD-SVD 823  
Db 425 ERTEQKTEVEETAGSVPAEELVGNDAEPQAEPAKELVKLTKETCVSGEDPTQGADLSPD 484  
Qy 824 VEV--RVYPENHASKVEGDNTKER-----DLDRASEKVEP-----RDELV 861  
Db 485 EKVLSPPEGVVSEVEMLSQSRMKVQGSPLKLTSTGLKLKSGKKQKGRGGDESG 544  
Qy 862 VAQINARPEPQSDNDSSATCSADE-----VQGEPERQ 896  
Db 545 EHTQVPADSPDSQEEQKGESSASSPEEPTITCLEKGLAEVQDGEAEAGATSDGEKKRE 604  
Qy 897 RMFPMDSKPSLLNPTGSLVSSPLKPNPLDLPQLQHRAAVIPPMVWVCTPCNIPITGVPSG 956  
Db 605 GVTWASFKKMVT----- 618  
Qy 957 YALYQRHIKAMHESALLEEQRQEQIDLECKRSTSPGTSKSPNREWEVLQAPHOLIT 1016  
Db 619 -----KKVRPPSES-----DKEDLDKVKASATLSSTESTASEMQEEMKGSVEEPKEE--- 667  
Qy 1017 NLPEGVRLPTTRTPRPPPLIPSSKTTVASEKPSFIMGSGISQGTGTYLTSHNQASYTQ 1076



Db 668 -----PKRKVDTSVSWEALICVSSKK-----RARRRSSDSE 699  
QY 1077 ETPKPSVGSISLGLPROQESAKSATPVYKOEFSRSONSQPGLLVRAQHEGVVRCGA 1136  
Db 700 EGGPKAMG-----DHQKADE-AGDKETGTDGILAGSQEHDPGGCSS 741  
QY 1137 GAIQEGSITRTPTSKISVESIPSLRGSITQGTALPQGTPTPTALVKGSGSRMPIDSS 1196  
Db 742 SPEQAGS-----PTEGEGVSWESFKRLVTPRKSKLEKSEDSIAGS-----GVEHST 792  
QY 1197 P--EKGREAAKSHVYEGKSHILSYDNKNAREGTRSPRTAHEISLKRYSVESVGN 1254  
Db 793 PDTEPGKEE-----SWVSITKFTIPGRKKRP-----DGKOEQAPVEDAG 831  
QY 1255 KQMSMRSPVSA--PL-----EGLICRALPRGSPHSDLKERTVLGSGIMQGTPRAT 1304  
Db 832 PTGANEDSDVPVAVVPLSEYDAVEREKWEAQAGAEQPKAATEVSKELSESQVHMM 891  
QY 1305 TESFEDGLYKPKIKRESP-----PIRAF--EGAITKGPYDGTITTIKEMGRSIE 1353  
Db 892 AAADVADGTRAATIIERSPSMISASVTEPLEQVEAEAAALLTEEVLEEVIAEEPPVTE 951  
QY 1354 -IP-----RQDILQESKTEVQVSTRPIIEGSIQGTPIKFNNSQCSAIKHN----- 1402  
Db 952 PLPENREARGDTVVSEAEITPEAVTAA-----ETAGPLGSEEGTEASAAEETTEMVS 1003  
QY 1403 -VKSITGPKSLSRGMPLEI---VPENIKVVERGKYEDVKA-GETVRSRHTSVVSSGP- 1456  
Db 1004 AVSQTDSPTTEATPVQVEGGVPD-IEQERTQEVLOQVAEKVEESQLPCTGQPE 1062  
QY 1457 SVLRSTLH---EAPKAOL-SPGIVDDT-----SARRTPVSYQNTMSRGP-MMNR 1502  
Db 1063 DVLPQVORAEAEPEQAEASGLKKTVDVLKVDQAQEAKEPFTQGVVGTTPSEFEKA 1122  
QY 1503 SDVT--IPPKNSTHNRKSTILTPQRESIPAKSPVPGVDPVVSHPDPHURGSTAGEVY 1560  
Db 1123 PQVTESIESELVTTCQAEITLGVKSQEMVNEQAIP---PDSVETPTDSETDGTPTVADF 1179  
QY 1561 WSHLPTQLDPAMPFHRA---LDPA---AAAYLFQRLSPTPGYPQYQLYAMENRTOT 1612  
Db 1180 DAPGTTQKDEIVEIHEENEVHLVPVGTAEAVPAQKERPA---PSF-VFOETREQS 1235  
QY 1613 ILNDYIYSQQMVLNRPDVARGLSPR-----EQPLGLP-YPATRGIIIDLNNPP 1660  
Db 1236 KWEDTLEHTDKEVSV--ETVSILSKTEGTQADQYADEKTKDVPFFEGLEGSID-TGI-- 1290  
QY 1661 TILVPHPGGTSTPMDRITYI---PGTQITFPR---PYNASMSRPHGHTHLAAASA 1712  
Db 1291 -----TVGREKVEVALKGEETEAECCKDDALELQSHAKSPSPVEREMVQV 1339  
QY 1713 ERER-ERE---REKERERIRIAASDLYLRPGSEQ-----PGRPGSHGYVRSP 1757  
Db 1340 EREKTEAETHVNEKLEHETAVTVSEV-----SKOLLQTVNVPILIDGAKEVSSLEGSP 1394  
QY 1758 SPVKTQETM-----LQORPSVFGTNGTSTVITPLDPTAQLRIMPL-----PAGGPSISQ 1807  
Db 1395 PPLCGLQEAECVTKIQVOSSEASFLLTAAAEKEKVLGETA--NILETGETLEPAGAHVLVE 1452  
QY 1808 GLPASRYNTAA---DALAALVDAASAPQMDYS-----KTKESKHAAR 1848  
Db 1453 EKSEKNEBFAHPGEDAVPTGPPCQAKSTPVIIVSATTKKGLSDDLGEKTTSLKWSDE 1512  
QY 1849 LEENLRSRAVS---BQOQLEQKTLVEKRS-----VOCLYTSSAFPSG 1890  
Db 1513 VDEQVACQEVKSVIAIELEPENILELETKSKLVQNIQTAVDQVFRTEETATEMLTS 1572  
QY 1891 KPQPHSSVVVSEAGKDGPPPKSRYEE--ELNRGKTTITAAANFIDVITTRQIASDKDARE 1949  
Db 1573 ELQTOAHVIKADS-QDAQETEKEGEPEQASAOETPITS----- 1612  
QY 1950 RGSOSSDSSSLSHRVTETPSDALEVISPASSPAPPOKEQLQTOYQEVVVKANQENDPTRQ 2009  
Db 1613 --KEESESTAVGOAH-----SDISKDMSEAS-----EKTMTVEVE-----GSTVND--QQ 1653

QY 2010 YEGPLHHRPQESPSQOOLPPSSQAEQMG--OVPRTHLRLITLADHICOLIITO----- 2061  
Db 1654 LE-----EVLVPSSEEGGAGTKSVPEDDGHALLAERIEKSLVPEKDEK 1698  
QY 2062 ----DFARNQVSS-----QTPOQPPTSTFONSPS-----ALVSTPVRTKTS 2098  
Db 1699 GDDVDDPENQNSALADTADASGLTKESPDT----NGPKQEKEDAQVEQLQEGKVHSESD 1754  
QY 2099 NRYSPESQASQVHHQRPGRS 2118  
Db 1755 KAITPQAEELQKQERESAK 1774  
RESULT 20  
US-08-194-468-2  
; Sequence 2, Application US/08194468  
; Patent No. 5750336  
; GENERAL INFORMATION:  
; APPLICANT: Montminy, Marc R.  
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN  
; TITLE OF INVENTION: RESPONSIVE GENES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/194,468  
; FILING DATE: 10-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9672  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619)-546-4737  
; TELEFAX: (619)-546-9392  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2441 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-194-468-2

Query Match 2.4%; Score 309.5; DB 1; Length 2441;  
Best Local Similarity 18.7%; Pred. No. 1.5e-10;  
Matches 495; Conservative 304; Mismatches 992; Indels 859; Gaps 122;  
QY 7 PPNQAGAFSTEQSRYPHPSVQYTFPNTRHQQEFAVPDYRSSHLEYSQASQLLQOQOQOQ-- 64  
Db 121 PLNQGDST-----PNLPKQA-----ASTSGPTPPASQALNPQAQKQV 159  
QY 65 LRRRPSLLSEFHGSDRQPRRTSYEFPHG--PSPVDHDSLESKRPRLEQVSDSHF--- 119  
Db 160 LVTSSPATSTQGTGIC-----MNAFNQTHPCLLNSNGHSLMNOAQOQOQVWNGSLGAA 215  
QY 120 QRVSAAVLPLVHPLPEGLRASADAKK---DPAFGGKHEA----- 155  
Db 216 GRGAGCMPTAPAMQCATSVLAETLTQVSPQWAG-HAGLNTAQAGMTKMGWTGTSP 274  
QY 156 ---PSSPISGQPCGDDQNASPSKLSKEELIQSDMRDVRDREIAKVBQQLKLLKKKQOOLEE 212



























Qy 1845 ---EAARLEENRRAAASVSEQQLEKLEVEKRSVQ-----CLYTSSA 1886  
Db 1406 FSVLSKSSNN-KLNLNAKKKQVQKAILQKNKSAKADLKNACSSSHICPYCNRE 1464  
Qy 1887 FP-SCKPQPHSSVYSEAGKDKGPPPKSRYEELRTRGKTTITANFDIVITRIASDK 1945  
Db 1465 FTYIGSLNKHAA--FSCPCKPLSPPK-----KVS 1492  
Qy 1946 DARERGSSDSSS---SLSSHRYETPSDAIEVISPASSPAPQEKLOTYOPEVVKANO 2002  
Db 1493 HSSKKGHSSPASSDKNSNHRRT-----ADAEIKMSMOTPLGKTRAR 1538  
Qy 2003 ENDPTRQVEGLHYRPOE---SPSPQOQLPPSSQAGMGQVPRTHLITLADHICQII 2059  
Db 1539 SSGPT-QVPLPSSSPRSQNVKFAASVSKKPSSSSLRNSPI-----RWAKIT-HVEGK 1592  
Qy 2060 TODFARN---QVSSOTPOOPTISTFONPSAL--VSTPVRTKTSNRYSPESQAQS 2109  
Db 1593 PRAVAKNHSQLSKTSRLHVRV--OKSAVLQSKSLASKKRTDRFNKRSRS 1646

## RESULT 25

US-08-046-585-5

; Sequence 5, Application US/08046585

; Patent No. 5453362

; GENERAL INFORMATION:

; APPLICANT: Lamarco, Kelly

; APPLICANT: Wilson, Angus

; APPLICANT: Hertz, Winship

; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

; TITLE OF INVENTION: HOST CELL FACTOR

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON &amp; HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/046,585

; FILING DATE: 12-APR-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman, Richard A

; REGISTRATION NUMBER: 36, 627

; REFERENCE/DOCKET NUMBER: A-57503-1/RAO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2035 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-046-585-5

Query Match 2.3%; Score 294.5; DB 1; Length 2035;  
Best Local Similarity 19.5%; Pred. No. 9.8e-10;  
Matches 412; Conservative 237; Mismatches 839; Indels 623; Gaps 98;

Qy 540 EKESDKNTKEKDKIDGTAET---EEREQATPRGRK-----TANSQGRKK 583

Db 287 EKWKCTNLACLNDTMAWETILMDTLEDNIPRARAGHCAVAINTRLYINSGRDGYRKA 346

Qy 584 RITRSMTEAAAAAASAAAAATEEPPPLPPPEPISTPVPVETSRWTEEMEVAKGLVEH 643  
Db 347 -----WNNQVCCCKDLWLETEKPPPP-----ARVOLVRANTNSLEV----- 382  
Qy 644 GRNNAIAIAKAVGTGKSEACKNFYFNKRRHNLNLLQHKOKTSKRPREERDVSCESA 703  
Db 383 ---SWGAVA-----TADSVLLQ-----LQKYDI---PATA 406  
Qy 704 STVSAQEDIEDIEASNEENPEDSEAVKPSDESPENATS--RGNTEPAVELEPTEAP 761  
Db 407 AT-----ATSPPTNPVPS--VPANPKSPAPAAAAPAVQPLTQVGTITLLPQAAP 455  
Qy 762 STPSLAYPTSKPADESQVNDISIAETAEDMDVQOEHSABEG-----SVCDDPPPA 816  
Db 456 PTTITQIVLPVP-----GSSISVPYA-----ARTQGVPAVLKVTGPQAT 495  
Qy 817 TKASVDVEVRVNPENHASKVEGDNTKERDLORASEKVEPRDEDLVAAQ-----INAORP 871  
Db 496 TGTPLVTMR---PASQAGKAPVTTS-----LPAGVRMVVPTQSAQGTIGSSP 541  
Qy 872 EPQSDNDSSATCSADEVDGEPERQRMFMDSKPSSLN-PTGSILVSS-PLKPNPLDLPO 929  
Db 542 QMSGMAALAAAAATQKI-----PPSAPTVLVSPAGTTIIVKTMVTPGTTILPA 591  
Qy 930 LQHRAAVIPPMVSVCTPCN-----IPIGTPVSGYA-LYQRHIKAMHESALLEEQRQOE 981  
Db 592 TV-KVASSPVMVS-NPATRMLKTAQAQVGTSSVSSATNTSTRPIITVHKSGITVVAQQAQ- 648  
Qy 982 QIDLECRSSTPCGTSKSPNREWEVLQAPAHQIINLPEGVRLPTRP----- 1029  
Db 649 VVTVVGGVTKTITLVKSP-----ISVPGGSALISNLKVMVSVQTKPVQTSVAVTQAST 703  
Qy 1030 -----TRPPPLIPSSK-TTVASEKPSFIMGSIQSGTPGTLYLTSHNQASYTQETPK 1080  
Db 704 GPVTQIIQTKGRLPAGTILKLVTSADGKPTIITITQASGA-GTKPTILIGISSVSPSTK 762  
Qy 1081 PSVGSISLGLPRQ-----QESAKSAT-LPYIKOEESP-----RSONSOPEGLL--- 1123  
Db 763 PGTITIKTIPMSAIIIOAGATGVTSSPGIK---SPITITTKVMTSGTGAPAKIITAV 818  
Qy 1124 ---VRAQH-----EGVVRGTAGAIQEGSITRGT-----TSKISVESI-PSLRSGITQG 1168  
Db 819 PKIATGHGQGGQVTVQVLKGAQG--QPGTILRTVPMGGVRLVTPVTVSAVKPAVTTLVKVG 876  
Qy 1169 TPALPQTCIPIEALVKSGISRMPIEDSSPEKREAAKSHVIEGKSHILSYDNIANA 1228  
Db 877 T-----TGVTTGLTGTGTVS-----TSLAGAGGHSTASLATPITLTGIATL 919  
Qy 1229 REGTRSPRTAHEISLRSYVESVEGNIKQGMSPVSPAPLEGLICRALPRG---SPHSD 1285  
Db 920 SSOVINP-TAITVSAAOITLTAAGGLTT-PITMQPVSPQVQVTLITA-PSGVEAQPVHD 976  
Qy 1286 LKERTVLGSGSIMQGTPTATTESFEDGLKYPKOIKR--ESPPIRAPEGAITKGPYDGTIT 1343  
Db 977 LPVSILASPTTEQPTATVTIADSGQGVQPGVTVLVCSNPPCETHETGTNTAT---TTV 1033  
Qy 1344 IKEMGRSIEHPRQ-DILTQESRKTPEVVQSTRPIEGSISQ--GTPIKFDNNSQSAIK 1400  
Db 1034 VANLGG---HPQPTQVQFVCDRQEAAAASLVTSTVGQONGSVVRVCSNPPCETHETGTNTA 1091  
Qy 1401 HNVKSLITGPSKLSRGMPPLEIVPENIKVVERKGYEDYKAGETVRSRHTSVSSGPSVL 1460  
Db 1092 TTATSNMAGQHCCSN--PPCE-----THET----- 1114  
Qy 1461 STLHEAPKAQLSPGIYDDTSARRTPVSYONTMSRSGPMNRTSDVT-----IPPNKSNH 1515  
Db 1115 GTTNTATTAMSSVGNHQRDARR-----ACAACTPAVIRISVATGALEAAQGSQOC 1167  
Qy 1516 ERKSTLPTQRESIPAKSPVGVDPDVVSHSPFDPHHRGSTRAGEVYVWSHLPTQLDPMPFH 1575  
Db 1168 TROT SATSTTMTVMATGAP-----CSAGPL-----LGPSP--- 1197



QY 1576 RALDPAAYLFRQLSP-----TPG-----YPSQOYLAMENTRQILNDYITSOQM 1624  
Db 1198 -AREPGRSFAV-QLAPLSSKVLSSPSIKDLPAGRSHAVSTAAMT----- 1243  
QY 1625 VNLRPDVARGLSPROPL--GLPYPATRGIDLTWNPTILVPHPGGT---STPMDRIT 1679  
Db 1244 ---RSSVGAG-BFRMAPVCEISLQGGSPSTVTVTAL-BALLCPSATVTQVCSNPPCE--T 1296  
QY 1680 YIPGQITFPFPPRYNASMSPGHP-----THLAAASA-----EREREREKE 1723  
Db 1297 HETGTTNTATTSNAGSAORVCNPPCETHETGTTTATATATSNNGTGOPEGQPPAGRP 1356  
QY 1724 RERERIAASDLYLRPGSEQRPGSHGYVRS-----PSPSVRTQETMLQORPSVFOGT 1778  
Db 1357 CETHQTTSTGTTMSVSGALLPDATSSHRTVESGLEVAAPSVTPQA----- 1403  
QY 1779 NCTSVITPLDPTAQLRIMPLPAGGFSISOGPLPASRYNTAADLAALVDAASAAPOMDYK 1838  
Db 1404 -GTALLAPP-PTQRCVSNP-----PCETHETGTTTATVTTSNMSNQ-DPPP 1448  
QY 1839 TRKSKHEAARLENRSAAVSEQQOLEQKTLVEKRSVOCLYTSSAPPGSKPOPHSSV 1898  
Db 1449 AASDGEV-----ESTQGSVNITSSATTTVTSSTLTTRAVTTQSTVPVPGSVPPPEL 1504  
QY 1899 VYSEAGDKGPPKRYEEELTRGKTTI--TAANFIDVILITRIASDKDARERQSDD 1956  
Db 1505 QVSPGPRQOLPP-----RQLQASASTALMGSAEVLASASOTPELPAAVDLSSSTCEPSSG 1558  
QY 1957 SSSLSHSHRYETPSDAIEVISPASSAPPOEKL-QTYQPEVVVKANQANDTRQYEGPLH 2015  
Db 1559 QESAGSA-----VVATVVVQP---PPPTQSEVDQLSPQELMBAQAQATTLMVTG--- 1606  
QY 2016 HYRPOQSPSQOQLPPSSOAGMGQVPRTHLITLADHICQIITQDFARNQVSSOTPOQ 2075  
Db 1607 -----LTPBELVATVAAEAQAATEAQAALA---IQAVLQ--AAQAVMGTGEP 1652  
QY 2076 PPTSTFQNS-----PSALVSTPVTRKTSNRYSPESQAQSVHHR 2114  
Db 1653 MDTSEAAATVTQAEGLHLSAEGQEQATTIPVLTQQLAALVQQQLQEAQAQQQHHL 1712  
QY 2115 PGRSVSP-----ENLVDKSRGSPGK-----SPERSHVSSEPEYETSP 2152  
Db 1713 PTEALAPADSLNDPAIESNCLNELAGTVPSTVALLPSTATESLAPSNFTFVAPQPVVVASP 1772  
QY 2153 PQVPVHVKQDLSLLLSORGAPAPQRNDARSPGISYILPSFTK-----L 2198  
Db 1773 AKL-----QAAATLTVANGIESLGVKPKDLPPLPPSKAPMKKENQWFDVGI 1818  
QY 2199 ENTSPMV-----KSKKOEI-----FR--KLNSGGGD-S 2224  
Db 1819 KGTNVVTHYFLPPDDAVPSDDDLGTVPDYNOLKKQELQPGTAYKFRVAGINACGRGPF 1878  
QY 2225 DMAAAQ-----PGTEIFNLPAVTHSGSVSRGHGSFADPASNGLIEDIIRKALMGSDOK 2278  
Db 1879 ELISAEKTLPGFPAGP-----AIKTSKPDGAHLTWEPSPVTS-----K 1919  
QY 2279 VEDHGVMSQPMGVVPGTANTSVTSGSTRREGDPSP-----HSGVCKPKL 2326  
Db 1920 ILEYSVLIAQSSOAGGELKSTPAQLAFMRVYCGFSPCLVQSSLSNAHIDYTTKPAI 1979  
QY 2327 ISKSNRSKSPDKIPQGGYLGTRPSSSVSHSEGDIYHRQTCGWAMEDRPSSTGSTQFPYN 2386  
Db 1980 IPRIAARNEK-----GY-----GPATQVR-----WLQETSKDSSGTPKANKR 2016  
QY 2387 PLTMRMLSTP 2397  
Db 2017 PMSSPEMKSAP 2027

RESULT 26

US-08-393-703-5

; Sequence 5, Application US/08393703

; Patent No. 5585239

GENERAL INFORMATION:  
APPLICANT: Lamarco, Kelly  
APPLICANT: Wilson, Angus  
APPLICANT: Herr, Winship  
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
TITLE OF INVENTION: HOST CELL FACTOR  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,703  
FILING DATE: 24-FEB-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57503-2/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-393-703-5

Query Match : 2.3%; Score 294.5; DB 1; Length 2035;  
Best Local Similarity 19.5%; Pred. No. 9.8e-10;  
Matches 412; Conservative 237; Mismatches 839; Indels 623; Gaps 98;

QY 540 EKESKENTKDKIDGTAET---EEREQATPRGRK-----TANSQGRKG 583  
Db 287 EKEWKCTNLACLDJTMWETILMDTLEDNIPRAGHCAVAINTRLYIWSGRDGYKRA 346  
QY 584 RITRSMTNEAAAAAATAEPPPLPPPEPTSTFETSRWTEEMEYAKKGLVEH 643  
Db 347 -----WNNQVCKDLWYLETEKPPPP-----ARVQLVRANTNSLEV----- 382  
QY 644 GRNMAAIKMYGTSEACKNFYFNKRRHLDNLLQHKOKTSKPREEDVSOCEVA 703  
Db 383 --SWGAVA-----TADSYLLO-----LQYDI---PATA 406  
QY 704 STVSAQEDDEDTEASNEENPEDSEVAVKPSDESDSPENATS--RGNTPEAVELETTETAP 761  
Db 407 AT-----ATSPPTNPVPS-VPANPKSPAAAAAPAVQPLITQVGIITLLPQAAPAP 455  
QY 762 STSPSLAVSPKPAEDESVEQTQVNDSSISAEATAEQMDVQQEHSABEG-----SVCDDPPPA 816  
Db 456 PTTTITQIVLPV-----GSSISVPTA-----ARTQGVPAVLKVTGTPQAT 495  
QY 817 TKASVDVEVRVPENHASKVBECDNTKERDLDRASEKVEPRDEDLVAAQ-----INAORP 871  
Db 496 TGTPLVTMR---PASQAGKAPVTVTS-----LPAGVRMVVPTQSQAGTVIGSSP 541  
QY 872 EPQSDNDSSATCSADEVDGEPERQRMFPMDSKPSLLN-PTGSILVSS-PLKPNPLDLPQ 929  
Db 542 QMSGMAALAAAAATQKT-----PPSSAPTVLVSPAGTTIVKTMVTPGTTTLPA 591







REFERENCE/DOCKET NUMBER: FP-57503-1/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US93-11721-5

Query Match 2.3%; Score 294.5; DB 5; Length 2035;

Best Local Similarity 19.5%; Pred. No. 9.8e-10;

Matches 412; Conservative 237; Mismatches 839; Indels 623; Gaps 98;

Qy 540 EKEDSKENTKEDKIDGTAET---EEREQATPRGRK-----TANSQGRKRG 583  
Dy 287 EKWKCTNTLACLDNTWAEITLMDLTEDNIPRARACHCAVAINTRLYIWSGRDGYRKA 346  
Qy 584 RITRSMNEAAAAAATAEPPPLPPPEPTEPETSRTWTEEMEYAKGLVEH 643  
Dy 347 ---WNNQVCKDLWLETEKPPPP-----ARVQLVRANTNSLEV----- 382  
Qy 644 GRNWAIAKMGVTKSEAOCKNFYKRRHNLNLLQHKQTSRKPREERDVSCESVA 703  
Dy 383 ---SWGAVA-----TADSYLLO-----LOKDYI---PATA 406  
Qy 704 STVSAQEDIEASNEENEPDSEVAVKPSDSDPENATS---RGNTPEPAVELEPTETAP 761  
Dy 407 AT-----ATSPFNPNVPS-VPANPKSPAPAAAPAVQPLTQVGIILLFOQAAP 455  
Qy 762 STSPSLAVPTKPAEDSVETQVNDNISAEATQMDVDQOEHSAAEE-----SVCDDPPPA 816  
Dy 456 PTTTITQVLPVTP-----GSSISVPTA-----ARTQGVPAVLKVTGPQAT 495  
Qy 817 TKADSVDEVVRPENHASKVECDNTERDLDRASEKVEPRDELVAQO-----INAQRP 871  
Dy 496 TGTPLVTR---PASQAKAPVTVTS-----LPAGVMVPTQSAQGVIGSSP 541  
Qy 872 EPQSDNSATSCADEVDGEPEROMPEPMDSKPSSLN-PTGSLIVSS-PLKPNPLDLQP 929  
Dy 542 QMSGMAALAAAAATQKI-----PPSSAPVLSVPAGTIVKTMATVPGTTLIPA 591  
Qy 930 LQHRAAVTPMVSCTPCN-----IPGTPVSGYA-LYQRIKAMHESALLEBORQRE 981  
Dy 592 TV-KVASSPVMVS-NPATRMLKTAQAQVGTSSVSSATNTSTRPIITVHKSGTVTVAAQQAQ- 648  
Qy 982 QIDLECRSTSPCGTSKSPNREWEVLQAPHQLTNLPEGVRLPTRP----- 1029  
Dy 649 VTTTVVGVTITILVKS-----ISVPGSALISNLGKVMVSVMQTRPVQTSVAVTQAST 703  
Qy 1030 -----TRPPPLIPSSK-TTVASEKPSFMIGSGISQGPGLYTLSHNOASVTOETPK 1080  
Dy 704 GPVTQIQTKGFLPAGTLKLTSDGKPTTIIITQASGA-GTKPIILGISSVSPSTTK 762  
Qy 1081 PSVGSISGLPRQ-----QESAKSAT-LPYIKQEBFSP-----RSONSOPEGLL--- 1123  
Dy 763 PGTTTIITIPMSALITQAGATGVTSSPGIK-----SPITIIITKVMTSMTGAPAKIITAV 818  
Qy 1124 ---VRAQH-----EGVVRGTAGATQEGSITRGP-----TSKISVESI-PSLRGSIQ 1168  
Dy 819 PKIATGHGQQGVTVQLKGAPG---OPGTILVPMGVGRVRLVPTVTSVAVKPAVTTLVYVK 876  
Qy 1169 TPALPQGTIPTEALVKSGISRMPIEDSSPEKGREAAASKGHVIEYEGSGHLLSDNFKNA 1228  
Dy 877 T-----TGVTILGTGTGVS-----TSLAGAGGHSATSASLATPTITLGTATL 919  
Qy 1229 REGTRSPRTAHEISLKRYSVEGNIKOGMSRMSPSVASPLEGLICRALPRG---SPHSD 1285

Dy 920 SSOVINP-TAITVSAAOITLTAAGGLTT-PTITMOPVSQPTQVTLITA-PSGVEAQPVDH 976  
Qy 1286 LKERTVLSGSIHQGTTPRATTESFEDGLKYPQKIKR--ESPPIRAFEGAITKPKYDGTIT 1343  
Dy 977 LPVILASPTTEQPTATVTIADSGOGDVQPGTIVLVCNSNPCEHETGTGTTNAT---TTV 1033  
Qy 1344 IKEMGRSIEHPRO--DILTQESRKTPEVWQSTRIEISIQ--GTPIKFNNSSQSAIK 1400  
Dy 1034 VANLGG--HPQPTQVQVCDRQEAASLVSTVQOQSGSVVVCNSNPCEHETGTGTTN 1091  
Qy 1401 HNVKSLITGPKSLRGMPPLEIVPENIKVWERGKYEDYKAGETVRSRHTSVSSGSPVLR 1460  
Dy 1092 TTATSNMAGQHCSN--PPCE-----THET----- 1114  
Qy 1461 STLHEAPKAQLSPGIYDDTSARRTPVSVQNTMSRSPMMNRTSDVT-----IPNKNSTNH 1515  
Dy 1115 GTTNTATTAMSGVNGANHQDARR-----ACAAGTFAVIRISVATGALEAAQSGSKSOCQ 1167  
Qy 1516 ERKSTLTPTQRESIPAKSPVPGVDPVWVSHSPFDPHHRGSTAGEVYVWSHLPTQLDPAMPFH 1575  
Dy 1168 TRQTSATSTTVMATGAP-----CSAGPL-----LGPSPM--- 1197  
Qy 1576 RALDPAAAAYLFORQLSP-----TPG-----YPSQVQLYAMENRTOTILNDYITSQOMQ 1624  
Dy 1198 -AREPGRGSPAFV-QLAPLSSKVRLLSPSINKDLPAGRHSAAVSTAAMT----- 1243  
Qy 1625 VNLRPDVARGLSPREQL--CLPYPATRGIIDLNMPTILVPHPGGT---STPPMDRIT 1679  
Dy 1244 ---RSSVGAG--EPRMAPVCELSQGGSPSTVTVTAL-EALLCPSATVTVQCSNPPCE--T 1296  
Qy 1680 YIPGTQITFPFPPRYNSASMSPGHP-----THLAAAAA-----EREREREKE 1723  
Dy 1297 HETGTTNTATTSNAGSAQRVCNSNPCEHETGTHTTATTSNGTGOPEGQPPAGRP 1356  
Qy 1724 RERERIAAASDLRLRPGSEQPRGSHGYVRS-----PSPSVRTQETMLQORPSVFQGT 1778  
Dy 1357 CETHQTSTGTMTSVSGALLPDATSSHRTVESGLEVAAPSVTPQA----- 1403  
Qy 1779 NCTSVIPLDPTAOLRIMPLPAGGPSISOGLPASRYNTAADAALAAALVDAASAAPOMDVSK 1838  
Dy 1404 -GTALLAPF-PTQRVCNSP-----PCETHGTGTHTTATTVTSNNSSNQ-DPPP 1448  
Qy 1839 TKESKHEAARLEENLRSRSAVSEQQOLEQKTEVERKRSVQCLYTSSAFPSGKQPQHSV 1898  
Dy 1449 AASQGEV---ESTQGDVNIITSSALITTVTSITLFRATVTVTQSPVPGPSVPPPEL 1504  
Qy 1899 VYSAGKDKGPPPKSRVEEELRTRGKTTI--TAANFDIVITROIADSKDAREGSSSD 1956  
Dy 1505 QVSPGPRQQLPP-----ROLQASASTALMGESAEVLSASQTPELPAAVDLSTGEPSSG 1558  
Qy 1957 SSSSLSHRYETPSDAIEVISPASSAPPQEKL-QTYQPEVVKANQAEENDPTROYEGPLH 2015  
Dy 1559 QESAGSA-----VVAIVVQPP---PPPTQSEVDQLSLPOELMAEAQAGTTTLMTVG--- 1606  
Qy 2016 HYRPOQESPSQOQLPPSSQAEQGMQVPRHRLITLADHICQIITQDFARNQVSSQTPQ 2075  
Dy 1607 -----LTPELAVTAAEAAAQAATEEAQALA---IQAVLO--AAQAVMTGTEP 1652  
Qy 2076 PPTSTFONS-----PSALVSTPVRTKTSNRYSPESQAQSVHHQR 2114  
Dy 1653 MDTSEAAATVTAELHLSAEGQEQATTIPVLVTOQELAAALVQOQQLQEAQAQOHHHL 1712  
Qy 2115 PGSRVSP-----ENLVKSRGSRPCK-----SPERSHVSSEYEPISP 2152  
Dy 1713 PTEALAPADSNDPAIESNCLNELAGVSTVALLPSTATESLAPSNITFVAPQPVVVASP 1772  
Qy 2153 PQVPEVHKODSLLLLSQRAEPAEQRNDARSQGISYLPISFFTK-----L 2198  
Dy 1773 AKL-----QAAATLEVANGIESLVGKPDLPPLPPPSKAPMKKENQWFDVGI 1818  
Qy 2199 ENTSPMV-----KSKQEI-----FR--KLNSGGGD-S 2224  
Dy 1819 KGTNMVMTHTYFLPDDAVPSDDDLGTVPDYNOLKQLQELQFGTAKFRVAGINACGRGPF 1878







Qy 1118 QPEGLVRAQHEGVVRGTAGAIQBSITRGTPTSKISVESIPSLRGSITOGTTPALPOTGI 1177  
Db 995 QQE---ERGQOE---REVARLTQE---RGRAQADLALEK---AARAEI-----1030  
Qy 1178 PTEALVKGISIRMPIDESPSPKREAAASKGHVITYEKGSHILSYDNKIKNAEGRTSPRT 1237  
Db 1031 --EMRLQNALNEQREFAT-----LQEAHALATEKEGD-----1063  
Qy 1238 AHEISLKRISVESGNIKQGMMSRESVSPAPLEGICRALPRGSPHSDLKERTVLGSGIM 1297  
Db 1064 -QELAKLUGLEAAQ--IKLEELROT-VKQLKEOLA-----KKEKHAASGGA 1107  
Qy 1298 QCTPRATESPEDGLKPKQIKRESPPTRAPEGAITKGKPYDITTIKMGERSITHEIPRQ 1357  
Db 1108 QSEAGRTEP-----TGPKLEALRAEVSK-----1131  
Qy 1358 DILQESKRKTEPVVQSTRIPIEGSISOGTPIKFDNNSGQSAIKHN--KSLITGPSKLSR 1415  
Db 1132 --LEQOCQKQEQADS-----LERSLEAERASRAERDSALETLQGLEKEAQELGHSQAL 1185  
Qy 1416 GMPLEIVPENIKVVERGKYED-----VKAGETVRSRHTSVVSSGPSVLKSTLHEAPKAQ 1470  
Db 1186 ASAQRELAFAFTKYVDHSHKSAEDENKAQVARGRQBAERKNSLISS-----1229  
Qy 1471 LSPGTYDDTSARRTPSVYONTMGRSGPMNRTSDVTTPPNKSTNHERKSTLTPTQRESIP 1530  
Db 1230 ---LEEVSS-----ILNR-----QVLEKEGESKELKRLVMAESEKS--1262  
Qy 1531 AKSPVGVDPVVSHPDPHHRGSTAGEVYVWSHLPTQLDPAMPFHRALDPAA-----AA 1584  
Db 1263 -----QKLESCACCRCRQRPATVPQLNAA 1287  
Qy 1585 YLFORQLSPTPGYPSQYQVYAMENRQITLNDYITSQOMQVNLDPDVARGLSPREQPLGL 1644  
Db 1288 LLCGRRCRAS--GREAEKORVASENLROELTSQAERAEELQELKAWQEK-FFQKEQALS-1344  
Qy 1645 PYPATRGITLIDTNMPTTLVPHPGGTSTPPMDRITYIPGTITPTPPRPNYSMSPGHPT 1704  
Db 1345 -----TLQLEH---TSTQAL-----VSELLPA--K 1364  
Qy 1705 HL-----AAAASREREREKERE-----RERIAAASDL-----YLRGSEQPGRGS 1750  
Db 1365 HLCQQLAQEAQAAAEKREELQESQKAGGLRAELLRQELGELILPLRQKVAEQDRTAQ 1424  
Qy 1751 HGYVRSPPSVRTQETMLQORPSVFGTNGTSTVITPLDPTAQLRIMPLPAGGPSISGLP 1810  
Db 1425 Q--LRAEKASTAEQLSMLKKAHGLIAEEN-----RGL-1454  
Qy 1811 ASRYNTAADALAAALVDAASAPQMDVSKTKRESKHEAARLEENLRSRAAVSEQQQLBQKT 1870  
Db 1455 GERANLGRQFLEVELDOA-----REKYVQELAAVRADAETRLAEVQREASTARE 1504  
Qy 1871 LEVEKRSVQCLYTTSAPPSPGKPPQSHSVVYSEAGKDGPPPKPSYBEELTRGKTTTAA 1930  
Db 1505 LEV-----MTAKYEGAKVVLIERQFQOE---RQKLTAQVE 1538  
Qy 1931 NFIDVITRQIASDKDARE-----RGQSQSDSSSLSSHRYETPDAIEVISPASSPAP 1984  
Db 1539 E-----LSKKLADSDQAKVQVQKLVQAGGESQEQEORF-----1575  
Qy 1985 PQEKLOTYQPEVVKANAENDPTRQYEGPLHHY--RPOQESPPQOQLPPSSQ-----2035  
Db 1576 -QAQNLQELQALSQKQEAHAYKLQMEKAKTHYDAKKQONQELQELQSLQELQKQENKEL 1634  
Qy 2036 ---NEGMGVPRTHRLTL-ADHICQIITQDF--ARNQVSSQTPQOQPTSTFQNSPSALV 2089  
Db 1635 RAEAEIRGHELQQAGLTKAEQTCRLTAQVRSLEAQVAHADQQLRDLGKQFQVATDALK 1694  
Qy 2090 STPVRTKTSNRSPESQAQSVHHQRPGRSVSPENLVKRSGRSGKSPERSHSSEPYEP 2149  
Db 1695 SREPOARFQDLDSIDSLDSCBETPLSITS-----KLPTQDGTSPVGEFASP 1744

Qy 2150 IS-----PPQV-----PVVHEKQ-----DSL-----LLLSQRGAEPAEORN-----2180  
Db 1745 ISORLPKVESLESLYTTPPARSOAPLESLSLSDGVFLDSGKTRRSARRTTQIINIT 1804  
Qy 2181 -----DARSPGSGISYLSFFFT-----KLENTSPMVYKSKQEIFRKLNSGGGSDMAAAQP 2231  
Db 1805 MTKKLDVDEEPPDSAN--SSFYSTRSAPASQASLRATSTQSLAR-LGSPDYCNSALLSL-P 1860  
Qy 2232 GTEIFNLPAVTTSGSVSRG-HSFADPASNLGLDIIRKALMGSFDDKVE-----DHGVV 2285  
Db 1861 GYR-----PTTRSARRSOAGVSSGAPPGRN-----SFTMGTCQDEPQLDDOWNRIAE 1908  
Qy 2286 MSOPMGVVVPGTANT-----SVVTGSTRREEDPSPHSGVCKPKLISKSNSRKS 2335  
Db 1909 LQQRNRYCPHLKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRAS 1955  
Qy 2336 KSPI-----PGQYGLGTERPSSVSVHSEGDYHROTPGNWAMEDRPSSTGSTO-FPVNPLTMR 2391  
Db 1956 MQPIQIAEGTGITTRQQRKRVSLPHQG-----PG-----TPESKKATSCFP-RPMTPR 2003  
Qy 2392 -----MLSSTPTPIACAPSAVNOAAPHQ 2415  
Db 2004 DRHEGRKQSTEAQKKAAPASTKQADRRQ 2032  
RESULT 29  
PCT-US93-06160-4  
; Sequence 4, Application PC/TUS9306160  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/06160  
; FILING DATE: 19930621  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESO, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US93-06160-4

Query Match 2.3%; Score 286.5; DB 5; Length 2101;  
Best Local Similarity 18.4%; Pred. No. 3.2e-09;  
Matches 480; Conservative 336; Mismatches 924; Indels 869; Gaps 110;  
Qy 52 QASQLLQOQQOQQOOL-----RRRPS-----LLS--EFHPGSDRPPQERTSYEPHPG 95  
Db 48 EGQQLKQPVSERLDFVCSFLQNRKHPSSPECLVSAQKVLGEGSELELAKWTMLLY---104







Db 1695 SREPAKPOLDSIDSLDSCCEGTPLSITS-----KLPRTPDGTSTVGPASP 1744  
QY 2150 IS---PPOV-----PVYHEKQ-----DSL---LLLSQCAEPAEORN----- 2180  
Db 1745 ISQRLPPKVESLESYFFTPIDPARSOAPLESLSLGDVFLDSGRKTRSAARRTTQIINIT 1804  
QY 2181 -----DARSPGSIYLPSTFF-----KLENTSPMVKSKQETFRKLNSGGGSDMAAAQ 2231  
Db 1805 MTKKLDVEEPSAN--SSFYTRAPASQASLRATSTSTQSLAR--LGSPDYGNSALLSL-P 1860  
QY 2232 GTEIFNLPAVTTSGSVSRG--HSPADPASNLGLEDIIRKALMGDFDDKVE-----DHGW 2285  
Db 1861 GYR-----PTTRSSARRSQAGVSSGAPPGRN-----SFTMGTCQDEPEOLDWNRIAE 1908  
QY 2286 MSQPMGVVPGTANT-----SVTSGTTRREEGDPPSHSGVCKPKLISKSNRKS 2335  
Db 1909 LOQRNRVCPPHLKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRAS 1955  
QY 2336 KSPI---PGQGYLGTERTSPSSVSHSEGDYHRQTPGNAWEDRPSSTGTSTQ--FPYNPLTMR 2391  
Db 1956 MQPIQIAEGTITTRQQRKRVSLPHQG-----PG-----TPESKATSCFP-RPMTPR 2003  
QY 2392 -----MLSSTPPTPIACAPSVAVQAAPHQ 2415  
Db 2004 DRHGRKQSTTEAOKKAAPASTKQADRRQ 2032

## RESULT 30

US-08-466-390-4  
; Sequence 4, Application US/08466390  
; Patent No. 5686562  
; GENERAL INFORMATION:  
; APPLICANT: TOUTKATLY, GARAH P  
; APPLICANT: LIDGARD, GAVEL M  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

Query Match 2.3%; Score 285.5; DB 1; Length 2101;  
Best Local Similarity 18.48; Pred. No. 3.7e-09;  
Matches 480; Conservative 336; Mismatches 924; Indels 869; Gaps 110;

QY 52 QASOLLOOQOQOOL-----RRRPS-----LLS--EFHPSGRDQPERRTSYEPHPG 95  
Db 48 EQQOILKQPVSRDLDFVCSFLQKNRKHPSPECLVSAQKVLGSELELAKTMLLY--- 104  
QY 96 PSPVDHDSLESKRPLBQVSDSHFQVSAVAVLPLVHPLPEGLRASAD----- 142  
Db 105 -----HSTMSKSPDWEQFEYKIQAEALAVILKFLVDHEDGLNLEDENFLQAPVST 159  
QY 143 -----AKDPAF-----GKHAPSPISGQCGD----- 167  
Db 160 CSSTPPEELSPSHQAKREIRFLELQKVASSSSGNFLSGSPAS--PMGDITLQTPQFMR 217  
QY 168 -----DONASPSKLSKE--ELIOSMDRVDRETAKEVQOQILKLLKQKQOOLEEAAKPE 218  
Db 218 RLKQLADESRNDELELAENRKLITEKDAQIAMQQRIDRLALLN--EKQAASPLE 274  
QY 219 PEKVPVPPVQEKHRSIVQIYYDENRKAEEAHKIFEGLGPKVELPLYNPQSDTKVYHEN 278  
Db 275 PKE-----LEELRDKNESILMRLHETLK-----OCODLKTESQ 308  
QY 279 I--KTQVMRKLLIFPKRRNHARKQRE-----QKICQRYDOLMEAWKVKVDRIENPRR 331  
Db 309 MDRKINQLSEENGDLSEKLRFAHLQLOLDALNELTEHSHKATQEWLEKQAKLEKLSA 368  
QY 332 KAKESKREYVEKQPEIRKQREOQERFQ--RVGORGAGLSATIAARSEHEISEIIDGLSEQ 390  
Db 369 ALQDKKCLE-----EKNEILQKLSLEHLSQLQDNPPQEKGEVLQVLQ 415  
QY 391 ENNEKQMRQLSV-----IPPMFDAQRRVKFINNGLMEDPMKVYKDRQPMNVWTD 442  
Db 416 ETLKQEAATLAANNTOQARVEMLETERGQOQAKLLAERGHFEE-----EKQQLSLITD 470  
QY 443 HEKEI-----FKQFIOHPNFGL-----IASYLERKSVPCVLYIYITKKNENYKALVR 492  
Db 471 LQSSISNLQAKKELEQASQAHGARLTAQVAS-----LTSELTTLNATQ 515  
QY 493 RNYGKRGRNQOIARPSQEEKVEKEEDKAEKTEKKEE-----KKDEEE 537  
Db 516 Q-----QOQELAGLQOAK--EKQALQATLQOQEQASQGLRHQVLEQLSSLSKQEQ 566  
QY 538 KDEKEDSKENTKCKIDGTAETETEERQATPRGRKTYANSQGRKRGTRITSMWNEAAAS 597  
Db 567 LKEVAEKQEQATRQ--DHAQQLATAAEERE--ASLRERDAALKQ-----LEALE 610  
QY 598 AAAAAATEEPPPLPPPEPISTPEVTSRWTEEMEVAKKGLVEHGRNMAAIAKWGVTK 657  
Db 611 KERAKEILEILOQLOQVANEARDSAQTSVTAQREKAELSRK--VEELOACVETARQOHE 668  
QY 658 SEAOCKNFYFNKRRHNLNLLQOHKQKTSRKPR--EERDVQ-----CESVASTVSAQ 709  
Db 669 AQQVA-----ELEQLRSEQQKATEKERVQAEKQLOEQLOALKESLKVTKGSL 718  
QY 710 EDEIEASN--EENPEDESEVAVKPS-----EDSPENATSRGNTPEPAVELEPTT 757  
Db 719 EEEKRAADALEEQORCISELKAETRSLVFQHKRERELEEEERAGRGLEARLLQ----- 774  
QY 758 ETAPSTPSLAVPSTKPAEDESVESTQVNDTSIAETQMDVDQOEHSAEESVCDPPPAT 817  
Db 775 -----GEAQAETEVE---LRRELAEM---AAQHTAES----- 801  
QY 818 KADSVDVVEVRVPENHASKVEGDNTKERDLDRASEKVEPRDEDLVVAQOINAQRPESQDN 877  
Db 802 -----ECEQLVKEVAWR---DGYEDSQEEAQAQYGAQFQEQMLTKECEKARQELQ--- 850  
QY 878 DSSATCSADEVDGEPERQRMFMDSKPSLLNPTGSLVSSPLKPNPLDLPLQLOHRAVI 937  
Db 851 -----EAKEKVAG-----IESH-----SELQISROQNKLAEL 877  
QY 938 PMVSCPTCNPICPIGTPTVSGYALYQRIHAKMHESALLBEORQORQOIDECKRSSSPGCTS 997  
Db 878 HANLA-----RALQOQVEKVAQKALADDLSTLOEKM-----AATS 913  
QY 998 KSPNREWEVLQAPAHQILITNLPBGVRLPTTRTPPPPLIPSSKTTVASEKPSIMGSI 1057



















SEQUENCE CHARACTERISTICS:  
LENGTH: 2101 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-924-4

Query Match 2.3%; Score 285.5; DB 2; Length 2101;  
Best Local Similarity 18.4%; Pred No. 3.7e-09;  
Matches 480; Conservative 336; Mismatches 94; Indels 869; Gaps 110;

QY 52 QASQLLOOQOQOQL-----RRRES-----LLS---EFHPGSDRPOERTSTYFFHPG 95  
DB 48 EQOQILKQVPSELDVFSFLQKNRKHPSPECLVSAQKVLGSGSELELAKWTMLLY--- 104  
QY 96 PSPVDHDSLEKRRPLEQVDSUSHFORVSAAVLPLVHPILPEGLRASAD----- 142  
DB 105 -----HSTMSSKSPRDWQFQYKIQAEALAVILKFVLDHEDGLNLEDLENFLQKAPVPST 159  
QY 143 -----AKKDPAF-----GGKHEAPSPISGQPCGD----- 167  
DB 160 CSSTFPEELSPSHQAKREIRFLEQKVASSSGNFFLSGSPAS---PMGDILQTPQFQMR 217  
QY 168 -----DONASPSKLSKE---ELIQSDMRVDREIAKVEQOILKLLKKQOQLEEEAAKPE 218  
DB 218 RLKQQLADERSNRDELELAENRKLLEKDAQIATMMQORIDRLALLN---EKQAASPLE 274  
QY 219 PEKVPSPVPVQKHRSIVQIYIDENRKAABAHKIFELGPKFVELPLYNQSDTKVYHEN 278  
DB 275 PKB-----LEELRDKNESITMLHETLK-----QCQDLKTERKSQ 308  
QY 279 1--KTNOVMRKKLILFFKRRNHARKORE---OKICORYDQLMBAWEKKYVDRIENPNRR 331  
DB 309 MDRKINQLSENGDLSFKLREFASHLQLODALNELTEHSAQTQEWLEKQAQLEKLSA 368  
QY 332 KAKESKTREYVEKOPPEIRKQREOQERFQ-RVGQRGAGLSATIAKSEHSEIIDLGLSEQ 390  
DB 369 ALQDQKCLE-----ERNEILQGLKLSLEHLSQLQDNPNPQKEGVLGDVLQL 415  
QY 391 ENNEKQMPQLSV-----IPPMFDEAQRVRFINMGLMEDPMKVYKORQFNWVTD 442  
DB 416 ETLQGEAATLANNTQLOARVEMLETERGQOQAKLLAERGHFEE-----EKQQLSSLITD 470  
QY 443 HEKET-----FKDFIQHPKNFGL-----IASYLERKSVPCVLVYIYITKKNENKALVR 492  
DB 471 LQSSISLSQAKEELEQAQAHGARTAQVAS-----LTSELTLNATIQ 515  
QY 493 RNYGRRGRNOQIARPSQEEKVEEKEEDKAETKEKKEE-----KKDEEE 537  
DB 516 Q-----QDQELAGLKQAK--EKQAQLAQTLOQOQASQGLRHQVLEQLSSSLKQKEBQ 566  
QY 538 KDEKEDSKENTKENDKIDGTAEEETEEROQATPRGKTANSQGRKGRITRSWNEAAAS 597  
DB 567 LKEVAEQEATQ--DHAQOLAATAAEERE--ASLREHDAALKQ-----LEALE 610  
QY 598 AAAAATEEPPLPPPPPEPTEPVSRTVEEMEVAKGLVEHGRNWAIAAKMGVTK 657  
DB 611 KEKAKEILQLOQLQVANEARDSAQTSTYQAQREKAELSRK--VEELQACVETARQEQHE 668  
QY 658 SEACQKNFYFNKRRHNDLNLQOHKQKTSRKRP--EERDVSQ-----CSVASSTVSAQ 709  
DB 669 AQAQVA-----EELQLRSBQKATEKERVAKQEKDQLOEQLOALKESLKVTKGSL 718  
QY 710 EDEDELEASN--EENPEPDESEVANKPS-----EDSPENATSRGNTPEPVELEPTT 757  
DB 719 EEEKRAADALEEQORCISELKAETRSRIVEQHKRERKELEERAGRKLEARLLQL----- 774  
QY 758 ETAPTSPSLVSPSTKPADESVEVQNDNISAEATAQMDVDQOEHSAEESVCDPPPAT 817  
DB 775 -----GEAQAQATEV---LRRELAENM---AAQHTAES----- 801  
QY 818 KADSDVDVEVRPENHASKVEGDNTKERDLDRASEKVEPREDLVVAQOINQAQRPQPSDN 877

DB 802 -----ECBOLYKEVAAWR---DGYEDSQOEERQAQYCAMQEQLMTLKEBCEKARQELQ--- 850  
QY 878 DSSATCSADEVDGEPERQRMFMDKPSLLNPTGSIILVSSPLKPNPLDLPLQOLHRAAVI 937  
DB 851 -----EAKKVAG-----IESH-----SELQISRQONKLAEL 877  
QY 938 PMVSTCTPCNIPITGTPVSGYALYQRIKAMHESALLEBQROQROEIDLECRSSTSPCCTS 997  
DB 878 HANLA-----RALQOQVEKEVRAQKLAADDLSILOEKM-----AATS 913  
QY 998 KSPNREWEVLQAPHQLITNLPEGVRLTPTTRPPPLIPSSKTTVASEKPSFIMGSI 1057  
DB 914 KEVARLETVRKAGQOQETASRELKVPARAGDRQPEWL----- 952  
QY 1058 SQGTPGTYLTSHNQASYTOETPKPSVGSISLGLPQOQSAKSATLPYIKQEEFSPRSQNS 1117  
DB 953 --EEQOGRFCSTQAAQAMEREAQOMGN-----ELERLRAALM-----ESQOQ 994  
QY 1118 QPEGLLVRAQHEGVVRGTAGAIQEGSITRGPTSKISIVESIPSLRGSITQGTPTALPQTGI 1177  
DB 995 QOE---ERGOQE---REVARLTQE---RGAQADLALEK--AARAEI----- 1030  
QY 1178 PTEALVKSGISRMPIEDSSPEKGREEAASKGHVYIEYGSGHILSYDNINKNAREGTRSPRT 1237  
DB 1031 --EMRLQNALNEQRVEFAT---LQEALAHALTEKEGKD----- 1063  
QY 1238 AHEISLKSYSVEGNIKQGMRESPIVSAPLEGLICRALPRGSPHSDLKERTVLSGSI 1297  
DB 1064 --OELAKRLGLEAAQ--IKELEELRQT-VKQLKEQLA-----KKEKEHAGSGGA 1107  
QY 1298 QGTPRATTFESFEDGLKYPKQIKRESPIRAFEAGATKGPYDGIITIKEMGRSIEHIPRQ 1357  
DB 1108 QSEAGRTEP-----TCPKLEALRAEYSK----- 1131  
QY 1358 DILTOESKRTPEVQOSTRPIIEGSIISQGTPIKFDNNSQCSAIAKHNV--KSLITGPSKLSR 1415  
DB 1132 --LEQOQKQEQEADS---LERSLEAERASRAERDSALETLQGOLEEKALQELHGSQAL 1185  
QY 1416 GMPPLIEIPENIKVVERGKYED---VKAGETVRSRHTSVVSSGPSVLRLSTLHEAPRAQ 1470  
DB 1186 ASAOBELAAFTKQVDHSDKAEDWKAQVARGQEAERKNSLISS----- 1229  
QY 1471 LSPGIYDITSAARRTPVSYQNTMSRSGSPMMNRITSDVTIPPNKSTNHKSTLTPTQRESIP 1530  
DB 1230 ---LEEVS-----ILNR---QVLEKEGESKELKRLVMAESEKS-- 1262  
QY 1531 AKSPVPGVDVPSVSHSPDPHHRGSTAGEVYWSHLPTQDLPAMPFHRALDPA--AA 1584  
DB 1263 -----TLQLEH---TSQAL-----VSELLPA--K 1364  
QY 1585 YLFQRLSPTPCYQOYLAMENTROTILNDYITSOQOVNLRDPDVARGLSPREQPLGL 1644  
DB 1288 LLCGRRCRAS--GREAEKQVASENLRQELTQOARAELEGOELKAWQEK--FPQEQALS- 1344  
QY 1645 PYPATRGIIIDLTNMPPTILVPHPGSTTPPMMDRITYIPGTQITTFPPRPNYSASMSGPHGT 1704  
DB 1345 -----TLQLEH---TSQAL-----VSELLPA--K 1364  
QY 1705 HL-----AAAAAEREREREKERE-----RERTAAASSDL-----YLRPGEQPGRCGS 1750  
DB 1365 HLCQQLQAEQAARAEKRRHELEQSKQAAGGLRAELLRQARELGELIPLRQKVAEQERTAQ 1424  
QY 1751 HGYVRSPPSVRTQETMLQORPSVFGTNGTSVITPLDPTAQLRIMPLPAGGPSISQGLP 1810  
DB 1425 Q--LRAEKASYAEQLSMLKKAHLAAEEN-----RGL- 1454  
QY 1811 ASRYNTAADALAALVDAASAPQMDVSKTESKKEAARLEENLRSRSAASVSOQOLEKT 1870  
DB 1455 GERANLGRQFLEVELDQA-----REKVVQELAAVRADAETRLAEVQBEAQSTARE 1504  
QY 1871 LEVERKSYQCLYTSTSAFPPSGKPQPHSSVYVSEAGKDKGPPPKPSRYEEELRTGKTTITAA 1930



Db 1505 LEV-----MTAKYEGAKVKVLEERQRFQEE---RQKLTAAQVE 1538  
Qy 1931 NFIDVITRIQIASDKDARE-----RGSQSDSSSLSSHHRYETPDSDAIEVISPASSAP 1984  
Db 1539 E-----LSKKLADSOASKVQOQKLKAVQAGGSQQAQRF----- 1575  
Qy 1985 POEKLOTQVPEVVKANQANDPTROYEGPLHY -RPOQESPPQOQLPPSSQ----- 2035  
Db 1576 -QAQLNELQAOLSKQEAHAYKQMEKAKHYDAKKAQOQELQELRLSLQOLKENKEL 1634  
Qy 2036 ---AEGMGQVPRHLITL-ADHICQIITQDF--ARNQVSSQTFQOQPTSTFTQNSPQALV 2089  
Db 1635 RAEAEGLGHELOQAGLTKAEQTCRHILTAQVRSLEAQVAHADQQLDLGLGKFQVATDALK 1694  
Qy 2090 STPVRTKTSNRYSPESQAQSVHQRGSRGSRVSPENLVDKSRGSRGKSPERSHVSSEVPEP 2149  
Db 1695 SREPOAKPQLDLSIDSLSCIEEGTPLSITS-----KLPTQDGTGTSVPGEPASP 1744  
Qy 2150 IS-----PPQV-----PVVHEKQ-----DSL-----LLLSQRGAEPAEQNR----- 2180  
Db 1745 ISQRLPPKVESLESILYFTPIPARSQAPLESLSIDSLGDVFLDSGRKTSARRRTTQIINIT 1804  
Qy 2181 -----DARSPGSIYLPSSFT-----KLENTSPMVKSKQETFRKLNSSGGGSDMAAAQ 2231  
Db 1805 MTKKLDVEEPSAN--SSFYSTRSAPASQASLRATSTQSLAR-LGSPDYGNALLSL-P 1860  
Qy 2232 GTEIFNLPAVTTSGSVSSRG-HSFADPASNLGLEDIRKALMGSFDDKVE-----DHGVV 2285  
Db 1861 GYR-----PTTSSARRSQAGVSSGAPGORN-----SFTMGTCQDPEQLDWNRIAE 1908  
Qy 2286 MSQPMGVVPGTANT-----SVVTSGETRREEDGDPSPHSGVCKPKLISKSNSRKS 2335  
Db 1909 LQQRNRVCPPLKTCYPLESRPSLSLGTITDEEMKTGDP-----QETLRRAS 1955  
Qy 2336 KSPI---PCQGYLTERPSSVSVHSEGDYHROTPGHNWEDRPSSTGSTQ-PPYNPLTMR 2391  
Db 1956 MOPIOAEGTGITTRQQRKRVSLPHQO-----PG-----TPESKATSCFP-RPMTPR 2003  
Qy 2392 -----MLSSSTPPTPIACAPSNAVQAAPHQ 2415  
Db 2004 DRHEGRKOSTTEAQKKAAPASTKQADRQ 2032

RESULT 34  
us-08-056-200-94  
; Sequence 94. Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/056,200  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: Fedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-056-200-94

Query Match 2.2%; Score 284; DB 1; Length 1898;  
Best Local Similarity 18.5%; Pred. No. 4e-09;

Matches 210; Conservative 170; Mismatches 376; Indels 382; Gaps 44;

Qy 33 RHQDEFVDPYRSSHLEVSQASQLQOQO---QOQLRRRPSLLSEFHPGSDRPOERTSY 89  
Db 397 RREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRRE 449  
Qy 90 EPFHPGPSVDHDSLESKRP-----RLEQVSDSHFQVSAAVLPLVHPLPEGLRAS 141  
Db 450 EERHEQKEQERREOQLRREOQERRDLWKREETERHEQE-----RRKQ 493  
Qy 142 DAKDPAFGGKHEAPSPISQPCGDDONASPSKLSKEELIQSDMDVDRETAKVQOILK 201  
Db 494 QLRDQE-----EERRRWLKEEERREOQERREOQLRREOQERRE 535  
Qy 202 LKKQOQLEEEAAAPPEKPPVPPVEQKHRSIVQIYIDENRKKAEAAHIFEGLGPKV 261  
Db 536 QRLKQEEERLQRLSEQOQLRR---EQEER-LEQLLKREERKRLQERR----- 582  
Qy 262 ELPLYNQPSDTKVYHENIKTNQVMKKLILFFKRRNHARKQ---REQKICORYDQLMQAW 318  
Db 583 -----EQLRKREOQERRDQLL--KREERROQLKREQE--ERLEQRLK-- 622  
Qy 319 EKKVDRIENPR---KAKES-----KTREYVEKQFPETKQREOQER----- 358  
Db 623 REEVERLEQEEERDRERLKRREEERHELKSEQERRHEQLR--REOQERREQLR 680  
Qy 359 ---FORVQORGAGLSATIAARSEHISEIIDGLSEQENNEKQMLSVPPMMF-----D 409  
Db 681 EEEERLEQR-----LKREHEERREQELAEQEQARERIKSRIPKQWQWQLESEAD 732  
Qy 410 AEQRRVKFINNGLMEDP-----MKVYKDRQFMVWTDHEKEIFKDKFTQHPKNF 459  
Db 733 ARQSKV-----LLEAPQAGRAEAPQEQEKRRRESELOQWEEERAHROQO----- 777  
Qy 460 GLIASYLERKSVDPCLVLYYLTKKNENYKALVRNRYKRRGRNQOIARPSQEKVE----- 515  
Db 778 -----EEQRDRFTWQWQAEKSE-----RGRQLRSARPLPQERERQLR 817  
Qy 516 -----EKEEDKAEKTEKKEEKK-----DEEKKDEKSKENTKEKDKIDGTA 558  
Db 818 AEERQOQERFLPEEEKEQGRORREKEXELOFLLEEEOQLORREQAQLOQEEE--DGLQ 875  
Qy 559 EETEERQATPRGRKTA-NSQGRKKGRITRMTNEAAAAAASAAAAATEEPPLPPPPPEP 617  
Db 876 EDQERRRQEQRRDQKWRQWLEERKRR-----RHTLYAKPALQEQLRKEQQLQEEEEE 929  
Qy 618 ISTEPTVTSRWTEEMEVAKKGLVEHGRNWAAIAKMGVGTKEAOCKNFYFNFKRRHLDN 677  
Db 930 LQREERKRRQEQERYREEEQLEQEE-----QLRREERKRRQO---EREQYRKDK 981  
Qy 678 LLQO-----HKQTSRKPREERDVSCQESVASTVSQAQEDIEDASNEENPE 724  
Db 982 KIQKQEQLLGEERPEKRRRQERREKKYREEEELQO-----EEOLLREERKRRQ 1031  
Qy 725 DSEVEAVKPSDSDPENATSRGNTEPAVELEPTTAPSTSPSLAVPSTKPAEDSV----- 780



Db 1032 EWEQYRKDELQEE-----EQLREERKRLQERQYREBELQEEELQEE 1084  
Qy 781 -ETQVNDISAEATQMDVDQOEHAEGSVCDPPATKAUSDVVEVPE----- 830  
Db 1085 RETRRQLEQYRKDELQEE-----EQLLEPEKRRQERQYREBELQEEELQEE 1140  
Qy 831 -----NHASKVEDNTKRDLDRASEKVEPRDELYVAQINNAQRP-----EPQSDN 877  
Db 1141 LREERKRRQLEQYREBELQYRKQRYRDE-----QRSDLKQWQPEKEN 1191  
Qy 878 ---DSSATCSADE-----DVGGEPRQRM-----F 899  
Db 1192 AVRDNKYCKRGNEQFQLEDQSVDRQSDQDQLHLLGEQERQYREBELQEEELQEE 1251  
Qy 900 PMDSKPSLLNPTGSLVSSPLKPNLDPQLQHRAAVIPPMVSCPTPCNIPGTVPVGYAL 959  
Db 1252 PEEQ-----LEREQKEA----- 1265  
Qy 960 YORHITKAMHESALL-----EQRQROEQIDLECRSSTSPCGTSKSPNREWEVLQAPAH 1013  
Db 1266 KRDRKSOEKQLLREERKRRQEE-----TDRKPREELQLEQEEQ 1309

## RESULT 35

US-08-800-644-94  
; Sequence 94, Application US/08800644  
; Patent No. 5958752  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichochalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,644  
; FILING DATE: 14-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/056,200  
; FILING DATE: 30-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fedrick, Michael F.  
; REGISTRATION NUMBER: 36,799  
; REFERENCE/DOCKET NUMBER: NIH054.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (714) 760-0404  
; TELEFAX: (714) 760-9502  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1898 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-800-644-94

Query Match 2.2%; Score 284; DB 2; Length 1898;  
Best Local Similarity 18.5%; Pred. No. 4e-09;

Matches 210; Conservative 170; Mismatches 376; Indels 382; Gaps 44:  
Qy 33 RHQEFAYPDYRSSHLEYSQASQLLQOOQ---QQQLRRRPSLLSEFHPGSDPPQRRSY 89  
Db 397 RREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRRE 449  
Qy 90 EPFHPGPSVDHDSLESKRP-----RLQVSDSHFQVSAAVLPLVHPLPEGLRASA 141  
Db 450 EERHEQKHEQERQERQERQERQERQERQERQERQERQERQERQERQERQERQERQER 493  
Qy 142 DAKDPAFGGKHEAPSSPSIQPGDDQNASPSKUSKELIQSDMDRVREIAKVBQILK 201  
Db 494 QLKRDQEE-----EERRERWLKLEEEERREQERQERQERQERQERQERQER 535  
Qy 202 LKKQOQLEEEAAKPPPEKPVSPPPVQKHRSIVQIYDENRKKAEAAHIFEGLGPKV 261  
Db 536 QRLKQEEERLQRLRSEQLR---EQEER-LQLLKREEKRELEQERR----- 582  
Qy 262 ELPLYNQPSDTKVYHENIKTNQVMRKILILFFKRHNHARKQ---REQICORYDQIMEAW 318  
Db 583 -----EQLKREQEERRDQLL--KREERRQQLKREQE--ERLEQLRK-- 622  
Qy 319 EKKVDRIENNPR---KAKES-----KTRYIEKQFPPEIRKQREQBR----- 358  
Db 623 REEVERLEQERDRERLKRPEEPERREHLLKSEQEERREHQLR--REQOERREQLRK 680  
Qy 359 --FORVGORGAGLSATIRSEHEISEIIDGLSEQENNEKOMQLSVIPPMWF-----D 409  
Db 681 EEEERLEQR-----LKREHEERREQLAEEOQAREIRKSRIPKQWQWLESEAD 732  
Qy 410 AEQRRVKFINMGLMEDP-----MKVYKDFQFNMVMTDHEKEIFKQFIQHKNF 459  
Db 733 ARQSKV-----LLEAPQAGRAEAPQEQEKKRRESELOQWEEERARHQO----- 777  
Qy 460 GLIASYLERKSVDPCLVYLYTKNNENYKALVRRNYGKRRGNNOIARPSQEEKVE----- 515  
Db 778 -----EEQRRDFTWQAEKSE-----RGRQLRSARPLRQERQERQLR 817  
Qy 516 -----EKEEDKAETKEKEBEK---DEBEKDEKESKENTKDKIDGTA 558  
Db 818 ABERQORQRLPEEEERQGRQREKEKQFTLEEEQQLRRERAAQOLQEE---DGLQ 875  
Qy 559 EETEEREQATPRGRKTA--NSQGRKRGITRSMTEAAAAAASAAAAAATPEPPPLPPPEP 617  
Db 876 EDQERRRQRRDQKRWQLEERKRR-----RHTLYAKPALQELRKEQOLQEEEEE 929  
Qy 618 ISTEPEVTSRWTEEMEVAKGLVEHGRNNAIAIKWGTSEAKQCNFYFNKRRHNDN 677  
Db 930 LQREERKRRQERQYREBELQEE-----QLLREERKRRRQ---EREROYRKDK 981  
Qy 678 LLQO-----HKQTSRKPREEROVSQESVASTVSQAQEDIEASNEENPE 724  
Db 982 KLOQKEQLLGEPEKRRRQREKRYREELQO-----EEQLLREERKRRRQ 1031  
Qy 725 DSEVEAVKPSDPSPENATSRGNTPEPAVELEPTTETAPSTPSLAVPSTKPAEDSV--- 780  
Db 1032 EWEQYRKDELQEE-----EQLLREERKRLQERQYREBELQEEELQEE 1084  
Qy 781 -ETQVNDISAEATQMDVDQOEHAEGSVCDPPATKAUSDVVEVPE----- 830  
Db 1085 RETRRQLEQYRKDELQEE-----EQLLREERKRRQERQYREBELQEEELQEE 1140  
Qy 831 -----NHASKVEDNTKRDLDRASEKVEPRDELYVAQINNAQRP-----EPQSDN 877  
Db 1141 LREERKRRQLEQYREBELQYRKQRYRDE-----QRSDLKQWQPEKEN 1191  
Qy 878 ---DSSATCSADE-----DVGGEPRQRM-----F 899  
Db 1192 AVRDNKYCKRGNEQFQLEDQSVDRQSDQDQLHLLGEQERQYREBELQEEELQEE 1251  
Qy 900 PMDSKPSLLNPTGSLVSSPLKPNLDPQLQHRAAVIPPMVSCPTPCNIPGTVPVGYAL 959  
Db 1252 PEEQ-----LEREQKEA----- 1265



Qy 960 YQRHKAHESALL-----EQRQROEQIDLECRSSTSCGTSKSPNREWEVLQAPHQ 1013  
Db 1266 KRDRKSOEKKQLREEREERROE-----TDRFREEQILLQREEQ 1309

RESULT 36  
US-08-227-536-2  
Sequence 2, Application US/08227536  
Patent No. 5658784  
GENERAL INFORMATION:  
APPLICANT: Eckner, Richard  
APPLICANT: Ewen, Mark  
APPLICANT: Livingston, David  
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/227,536  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Ph.D., Kathleen A.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: DFCI-308XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2414 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-227-536-2

Query Match 2.2%; Score 284; DB 1; Length 2414;  
Best Local Similarity 19.0%; Pred. No. 5.6e-09;  
Matches 454; Conservative 300; Mismatches 864; Indels 774; Gaps 118;

Qy 117 SHFQVSAVPLVHPLPEGLRASADAKDPAGGKHEAPSSPGCGDDQNASPSKL 176  
Db 401 SHWNCTRHDCVCLPLK---NAGDKRQOBILTGAPVGLGNPSS---LGVGQDSAPNLS 454  
Qy 177 SKEELIQMDRVREIAKV-----EQILKLLKKQOOLEEAAKPEPEKPSVSP 226  
Db 455 TVSQIDPS--SIERAYALGLPYQVQNMPTQVQAKNQOQCGSQDQGRPMNSNAS 512  
Qy 227 PVEQKHRSIVQ-----TIYDENRKKAEAKHIFPGLGPKVELPLYNQPSDT-- 272  
Db 513 PMGVNGGVGVQTPSLSDSLMSHAINSONPMWSENAS--VPSLGP---MPTAAQPSITGI 567  
Qy 273 -KVYHENIK---TNOVVRKKLILFFKRRNHARKOREKICORYDQIME---AWKKV--D 323  
Db 568 RKQWHEITQDLRNLHVLHVKLQVAFPTPDPAALK-----DRRMENLVAYARKVEGD 618  
Qy 324 RIENPRKAKESKREYI-----EKQFPEIKRQEQERF----- 359  
Db 619 MYESANNR-----AEYHLLAEKIYKIQELEKRRTRLOKQNMPLNPAAGVPSMNP 671

Qy 360 -QRYGQAGLSATISARSEHEI-----SEIDGLSEQEN-----NEKOMROLSVI---P 404  
Db 672 GPNMGQPOPGMTSNGPLDPSPMIRGSPVQNMWPRITQSGNLQFGQMSMAOPPIVPROTP 731  
Qy 405 PMMFAEQRVRKVFIMNGLMEDPMKVYKROFMVWTDHEKEIFKDFIQHPKPNGLIAS 464  
Db 732 PLOHHGQ-----LAQPCALNPPMG-YGPR-----MQQPSNOG---Q 763  
Qy 465 YLERKSVDPDCVLYYYLLKKNENYKALVRRNYGKRRGNQ---QIARPSOEKVEKEEDK 521  
Db 764 FLPTQTFPS-----OGMNVNIPLAPSSGQAPVSOAQM-- 796  
Qy 522 AEKTEKKEEKKDEEKKEDSKENTKEKDKIDCTAEETEEREQATPRGKTANSQRR 581  
Db 797 -----SSSSCPVNSPIPPG-----SQGSH 816  
Qy 582 KGRITRSMWNEAAAAAATAATEPPPLPPP-----PEPSTPEVETSRMT 629  
Db 817 IH--CPQLPOPALHONSPSPVPSRTPTPHPTPPSIGAQOQPPATIPAPVPTPP----- 867  
Qy 630 EEMEVAKKGLVEHGRNWAAIAKMGVTKSEAOCKNFYENKRRHNLNLLQOHKQKTSRK 689  
Db 868 -----AMPPGQSOA-----LHPPRQPTTP 888  
Qy 690 PREERDVSQCESVASTVSAODEDIEASNEEENPEDSEVAVK-PSEDS-----PENATSR 744  
Db 889 PTTQLPQVQVPSLPAAPSADQ-----PQQPRSQQSTAASVPTNAPLLPQPPATPL 940  
Qy 745 GNTPEPAVELEPTTETAPSTSP-----SLAVSTKPAEDESIVTQVNDISIAETAQMDVQ 800  
Db 941 --SQPAVSIEGQVSNPPTSTSTEVNSQAIAEKPSOEKMEA-----KMEYDQ 986  
Qy 801 QEHSAAEGSVCDPPPATKADSDVDEVVRPENHASKVEDNTERDLDRASEKVEPRUDEL 860  
Db 987 PE-----PADTQPEDIS-ESKVED---CKMESTETEERSTELKTIKEEDQPS 1031  
Qy 861 VVAQOINAAQPEPOSNDNS-----SATCSA--DEDVGEPEPQRMPMDSKPSL 907  
Db 1032 TSATQ---SSPAGQSKKIKFKPEELRQALMPTLEALYRQDPESLPRFQ---PVD--PQL 1083  
Qy 908 LN-PTGSLIYSSPLK---PNPLDLQLOHRAAVIPPMVSTPCNIPGTIPVSGVYALYQR 962  
Db 1084 LGIPDYFDIVKSPMDLSTIKRKLDTGQOEPWQYVD-----DIWLMFNANWLN 1133  
Qy 963 HIKAMHE--SALLEEQROBQIDLECRSSTSPGTSKSPNREWEVLQAPHLITNLPE 1020  
Db 1134 KTSRVKYCKLSEVFEQ---EIDPVMSGLYCCGRK-----LEFSQTLCC---Y 1178  
Qy 1021 GVRLPTRTRTPRPPPLIPSSKTTVASEKPSFIMGSSISQGTGTYLTSNHOASTQETPK 1080  
Db 1179 GKQLCT-----IPRD-----ATVYSYQNRHYHCEKCFN 1206  
Qy 1081 PSVG-SISLG-LPQOESAKSATLPIKOEFEPSRSONSQPEGLLVRA----- 1126  
Db 1207 EIQGESVSLGDDPSQPTT-----INKEQFSKRKNNDLDPPELFVECTEGRKMHQICV 1259  
Qy 1127 -QHEGVVGTAGTAQEGSITRGTP---SKLSVESIPSLR-GSITQG-----TPALPQ 1174  
Db 1260 LHHE--IIPWAGFVCDCLCKSAATRKENKFSAKRLSTRGLTLENRVNDFLRQNHPE 1317  
Qy 1175 TGITEALVKGSISSMPIE-----DSSPEKGREBAASKGHVIEYCKSHIISYNQIK 1226  
Db 1318 SGEVTVRVHASDKTVEVKPGMKARFVDSGEMAESFFYRTKALFAFEIDGVDLCFFGMH 1377  
Qy 1227 NAREGTRSP-----RTA--HEISLKSYESVEGNIKOGMSMR 1261  
Db 1378 VQEGSDCPPPNQRRVYISYLDSSVHFPRKCLRTAVVHEILI--GY--LEVYKKLGYTTG 1433  
Qy 1262 ESPVSAPLEG--LTCRALPRGSPHSLDKERTVLSGSIQMGTPPRATTE-----SFDGLKYP 1315  
Db 1434 HIWACPFSEGGDDYIFTHCHP---PDQKIPKPKRLQEWYKMKLDKAVSERIVHDYDKIF-- 1488  
Qy 1316 KQIKRESPPIRAFEGAITKGK--PY-DGIIITIKEMGRSIIHEIPRODILITQESRKTPEVQ 1372



```
Db 1489 -----QATEDRLTSAKELFYFEGDFWPNLEESIKELEQE-----EERKRE----- 1530
Qy 1373 STRPIEISQGTPIKFDNNSGOSAKHNKSLITGPKSLRGMPPLLEIYPENIKVVER 1432
Db 1531 -----ENTSNESTDV---TKGDSNAKKNNKNTSKNKSLSRGNKKKPGMPNVSNDLSQ 1582
Qy 1433 GYEDVKAGETVRSRHTSVVSSGSPVSLSTLHEAPKAQLSPGIYDD----- 1478
Db 1583 KLYATME-----KHKEVF---FVIR--LIAGPAANSLPPIVDPDLIPCDLMDGRDA 1629
Qy 1479 --TSARRTPSVYQ-----NTM-----SRGSPMMNRTS 1503
Db 1630 FLTLARDKHLEFSSLLRAQWSTMCMMLVELHTQSQDRFYVTCNECKHHVETRWHTCTVEDY 1689
Qy 1504 DVTIPPNNKSTHNEK-----STLTP--TQRESIP----- 1530
Db 1690 DICITCYNTKNDHMKMEKLGGLGDDSNQQAATQSPGDSRRLSIQRCISLVHACOCR 1749
Qy 1531 -AKSPVPQ---VDPVSVSHSPDPHHRGSTAGEVYWSHLPTQLDPAMPHPHALDPAAAAYL 1586
Db 1750 NANCSLPSCQMKRVQHT-----KCKRKTNGCGPICKQJLIALCCVH-----AKH 1795
Qy 1587 FOROLSPTPGVPSQYLYAMENTROTILNDVITSOQOVNLRPDVA-----RGLSPREOPL 1642
Db 1796 CQENKCPVP-----FCLNIKOKLROQLOHRLQQAOM---LRRRMAWMORTGVVGGQOGL 1847
Qy 1643 GLPYPATRGIIDLNMPTTILVHPGGTS---TPMDRITYIFGT-----Q 1695
Db 1848 PSPTPAT--PTTPTGOOPTTQTPQP--TSQPPTPPNSMPPYLPRTQAAQVPSQGAAGQ 1904
Qy 1686 ITFPPRPYNSASMPG-HPTHLAAAASAERERERERERERERERERERERERERERERER 1744
Db 1905 VTPPTPOTAPPLPGPPPTAVEMAQIQAETQO-----MAHVOIFORP----- 1951
Qy 1745 PGRPGSHGVYRSPSVKVTOTMLQQRPSVFGTNGTSVITP-LDPTAQLRIMPLPAGGR 1803
Db 1952 -----IQHQPMPPTMAPGMNPPMTGRPSGH--LEPGMGPTGMQOQPPWQSQGL 2000
Qy 1804 SISQGLPASRYNTAADALAAALVDAASAP-----QMDVSKTE---SKHEAARLEENLRS 1855
Db 2001 POPOLOSGMPPRPAMWSVAOHGQPLNMAPQGLQGVGSLPKGTVSQQAQLNLLTLRS 2060
Qy 1856 RSAVSEOOQL-----EOKTLEVERKSVQCLYTSSAFPSGKP--OP----- 1894
Db 2061 PSSPLOQOOVLISLHANPOLLAAFIKORAAKYANSNPQIPQCGMGQCGQLQPPPTWPG 2120
Qy 1895 -----HSSVYVS---EAGKDKGPPPKRSRYEEELR-----TRGKTTITAA NF 1932
Db 2121 QQGVHSNPAMQNMNPMQAGVQRAGLPQOQPOQOQLOPPMGMSPOAQQMNNHNTM-PSQF 2179
Qy 1933 IDVITITQIASDKARBERGSSOSSSSLSHRVETPDSAEIVSPASSPAPPOEKLOTY 1992
Db 2180 RDILRLOQMOMQOQ--OOQAGGIGPGMANHNPQOQOQ-----VGYPPOQOQRMQHH 2230
Qy 1993 QPEVVYKANAQNDPTROYEG-----PLHRYRP---QOESPSPOQOLPPSSQAGMGQVPR 2044
Db 2231 MQOQOQNGMOIGOLPQALGAEAGASQAYOQRLLOQOMGSPVQPNMS-----PQ 2281
Qy 2045 THRLITLAD--HICQIITQDFARNQVSQTPQOQPTSTFQNSALYSTPVRTKTSNRY 2102
Db 2282 QHMLPNQAQSHLQGOQLOIPNLSNQRS--POVPSPRQSQOPPH--SSP-----S 2328
Qy 2103 PESQAO-SVHH-----QRPGRSVSPENLVKSRGSRPGKSPERSHVSSEP 2146
Db 2329 PRMQPQSPHRSVQTSPPHGLVAAQANPMEOGHFASPDQNSMLSLASNP 2380
```

RESULT 37

PCT-US95-04682-2

; Sequence 2, Application PC/TUS9504682

; GENERAL INFORMATION:

; APPLICANT:

```
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,536
; FILING DATE: 14-April-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Heine, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-308Xq999
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04682-2
```

Query Match 2.2%; Score 284; DB 5; Length 2414;

Best Local Similarity 19.0%; Pred. No. 5.6e-09;

Matches 454; Conservative 300; Mismatches 864; Indels 774; Gaps 118;

```
Qy 117 SHFQVSAVPLVHPLEGLRASADAKDPAGFKGHEAPSSIPSGQPCGDDONASPSKL 176
Db 401 SHWNCTRHDCPVCLPLK---NAGDKRNQQLPILGAPVGLGNPSS---LGVQQQAPNLS 454
Qy 177 SKEELIOSMDRVDREIAKV-----EQOILKLLKKQOQLEEEAAKPEPEKPVSP 226
Db 455 TVSQIDPS--SIERAYALGLPYQVQNMPTQPOVQAKNQOQPCGSPQGRPMNSMAS 512
Qy 227 PVEOKHRSIVO-----IYDENRKAEBAAHIFEGLPKVELPYNQPSDT-- 272
Db 513 PMGVNGGVGVQVTPSLLSDSLHLSAINSONPMNSNAS--VPSLGP---MPTAAQSTTGI 567
Qy 273 -KVYHENIK---TQVMKKLILFEKRNHARKQEQKICQYDQLM-----AWKKV--D 323
Db 568 RQKHEDITQDLRNLHVHLVQALFPPTDPAALK-----DRRMENLVAYARKVEGD 618
Qy 324 RIENPRKAKESKREYV-----EKQFPEIRKQREQQERF----- 359
Db 619 MYESANNR-----AEYHLLAEKIYKQLEKRRTRLOKQNMPLNAAAGWVPSMNP 671
Qy 360 -ORVQQRGAGLSATIASEHEI-----SEIDGLSEQEN-----NEKQMLSVI----P 404
Db 672 GPNMGQPOPGMTSNGPLDPDPSMIRGSPVNMPPRTTQSGNLQFQCMSMAQPIVPROTP 731
Qy 405 PMFEDAQRVRKFINMNGLMEDPMKVYKDRQFMVWTDHEKEIFKQFIQHPKFNGLIAS 464
Db 732 PLQHHGQ-----LAQPGALNPPMG--YGPR-----MQQPSNOG---Q 763
Qy 465 YLERKSVDPDCVLVYYLTTKKNENKALVRNRYGKRGRNQ---QIARPSQEEKVEKEEDK 521
```







COUNTRY: U.S.A.  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/853,913  
FILING DATE: 19920319  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/660,412  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,803  
FILING DATE: 25-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/201,762  
FILING DATE: 02-JUN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/180,548  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1618 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-853-913-4

Query Match 2.28; Score 276.5; DB 1; Length 1618;  
Best Local Similarity 19.58; Pred. No. 9.3e-09;  
Matches 316; Conservative 215; Mismatches 521; Indels 565; Gaps 83;

Qy 10 QGAFSTEGSRYPHPSVQVTFNTRHQOEFAVPDYRSSHLEYSQASQLLQOQQOQLRRRP 69  
Db 186 RGAVRGYQERVA--HMETSLDQTRERLARAVOGAREVRLQO--LQAEKGGLLERRA 239  
Qy 70 SLLSEFHGSDRPRQERTSYEPFHGSPVDHD----- 102  
Db 240 ALEQRL-----EGRWQERLATEKFQLAVEALEQERKQGLSQIAQVLEGRQQLAHLKMSLSL 296  
Qy 103 -----SLESKRPRLE-QVSDSHFORVSAVLPLVHP--LP 134  
Db 297 EVAYRTILLEENSLQTPGGGSKTSLSFQPKLEQPPRTPEGRRLGSLPLVLSPTSLP 356  
Qy 135 EGLRASADAKDPAPGKHE-----APSSPIGQPCGDDQNASPSKLSKEELIQSMD 186  
Db 357 SPLPATLETPV-PALFKNQEFQARTPTLASTPIPTP-----QAPSPA----- 399  
Qy 187 RVDREIAKVEQIQLK-----KKQOOLEEA-AKP-----PEPEKP-----VS 224  
Db 400 -VDAEIRAQADAPLSLLQQTQGRKQAPLEPLRAEAVIPASVLPGPPEPGQORQEAQTQS 458  
Qy 225 P-----PPVEQKH-----RSI 235  
Db 459 PEDHASLAPPLSPDHSLEAKDGSGSRVFSICRGEQGIWGLVEKETAIEGKVYSSL 518  
Qy 236 VOIYDE---NRKAAEAH-----KIFELGPKVELPLYNQPSDTKVYHENIKTNQVMRK 287  
Db 519 QOEIWEEDLARKETQDSQVPLEKETLSLGEIQ-----ESLKT----- 558  
Qy 288 KLILFFRRNRHARKQREKICQRY-----DQLEWEAKVKVDRI-----ENNPRKAKES 336  
Db 559 -----LENQSHETLERENQECPRLEEDLETLSLEKENKRAIRKCGGSETSRKRCRQL 613

Qy 337 K-----TREYVEKQFPEIKKORE-QOERFORVQORGAGLSATIARSEHSEISIDGL 387  
Db 614 KPTGKEDTQTLOSLOKENQELMKSLGNLETFLPFGTENQELVSSLOENLESITAL----- 669  
Qy 388 SEQENNEKOMRQLSVIPPMFMDAEORRYKFINMGLMEDPMKVYKDRQFMNVTWTOHEKEI 447  
Db 670 -EKENQE-----PLRSPEVGDDEALR-----PLTKENQEPRLSRLE-----DENKEA 709  
Qy 448 FKDFIQHPKFNGLIASYLERKSVDCVLYYYLTKN-ENYKALVRYNRYKRRGRNQQA 506  
Db 710 FRS-----LEKENEQPLKL-----EEDQDSIV 732  
Qy 507 RPSOE-----KVEKEEDKAETKEKEEKKDEKEDSKENTKEDKIDGTAAE 560  
Db 733 RPLETENHKSLSRLEEQQOQETRLTLEKETQQR-----RSLGEQDQMTLRPPKVD--LEP 786  
Qy 561 TEEREQATPRGRKTANSQGRKGRITRSMTNEAAAA-----SAAAAATEPPPPPL 611  
Db 787 LKSLDQETARPLENENQE-----FLKSLKEESVAVKSLTEILESLKSAGQEN----- 835  
Qy 612 PPPPEPISTEPVETSRWTEEM-----EVAKKG-----LVEHGRNWAATA 651  
Db 836 ---LETLSKETQAPLWTPPEINKSGNERRKGNRTTGVCGSEPRDIOTPGRGESGII 892  
Qy 652 KMGVTKSEAOCKNFYFNKR-----RHNL---DNLLQOHKQKTSRKPREE-----RD 695  
Db 893 EISGSMERGE-----FEISRGVDKESQNRLEEEENLKGQEQESLSRLEEQQELPQSD 947  
Qy 696 VSQCESVASTVSAQEDDI-----EA-----SNEEENPEDEVEEA 730  
Db 948 VORWED-----TVEKDQELAQESPPGMAGVENKDEAELNLRQDQGTGKEEVVEQELNA 1002  
Qy 731 VK-----PSDESPEN-----ATSRGNTEPAVELETTETATPTSPSLAVPSTKP 774  
Db 1003 TEEVMFPGEGHPENPEKPEQRGLVEGASVKGGAEL--QDPEGQQQVGTGTLQAPQGLP 1060  
Qy 775 AEDSEVETOVNDSISA---ETAQOMDVQOEHSABE-----GSVCDPPPPATKA 819  
Db 1061 ---EAIPELVEDDVAPGDQASPEVNLGSEFAMGSAAGAPGLQGVGGGLDGPCHLIRE 1117  
Qy 830 DSDVDEVVRPNHASKVGDNTKERDLDRASEKVEPRDEDLVVAQQINAQR--EPQSDN 877  
Db 1118 EVMPPEPLEESLEAKRVQGLEGRPKDLEAGLGTEFSE--LPGKSROPWEPREGRES 1175  
Qy 878 DSSATCSADEVDGEPERQRMFPMDSKPSLNPCTGSLVSSPLKPNPLDLQLQRAAVI 937  
Db 1176 BAAAPRGABE-----APPAET-----LGTGSD-----AFSPWPLGSEEAEDVP 1215  
Qy 938 PMVSVCTPCNIPi-----GTPVSGVALYORHIKAMHES-----ALL 973  
Db 1216 PVLVSPSTYPTIILEDAPGLQPAEGSQEASGWQGRABEAGKVESEQELSGEIPGLO 1275  
Qy 974 EEOQRQEQIDLECRSSTSPCGTS-----KSP-NREWEVLQAPHLITNLPEGVRLPTT 1027  
Db 1276 EGEESRESEDEELGTLPDSTPLGFLYLRSPTSRPTWPLESRGHPKLTKEGK-- 1330  
Qy 1028 RPTRPPLPIPSKTTVASEKPSFMG-----GSISO-----GTPGYTLTSHNQASYT 1075  
Db 1331 ---DPAVLASBGLPEPSEKEEGEEGECROSDLSEEFDLGTEAPFL-----PGVP 1380  
Qy 1076 QETPKPSVGSISLGL-----PROQESAKSATLPYIKOEFSRPNQSQPEGLLVRAQHE 1129  
Db 1381 GEVAEP-LGQVPPQLLLDPAANDRDGED-----GFADSEESGEEGEEQEEG--REP 1429  
Qy 1130 GVVR-----GTAGAIOE-GSITRGPTTSKISVE-SIP-----SLRGSTQGTPTALPQTGPI 1180  
Db 1430 GAGRWGPGSSVGSILQALSSQRFLESDSVSVVPWDDSLRGAVA-----GAPKTALETE 1485  
Qy 1181 ---ALVKGSI SRM-PI-----EDSSP-----EKGREEAASKGHVI-----YEGKSGH 1218  
Db 1486 SQDSAEPGSGSEESDPVSLEREDKVPGLPTEIPSGMEDAGPGADIIGVNGQGNLEKSGQH 1545







Db 1909 TKLEPSEDDGKPELLEEMEAASPTLIAVEGTEILODFONKTDGOVSGEAIKMFPTIKTPE 1968  
QY 2199 ENTSPWAKKOEIIFKLNSSGGSDMAAQPGE-----INLPAVTTSGSVS 2248  
Db 1969 AGT---VITTADEIELE-GATQMPHSTASATYGVVWVWLSQTSERPTLSSSPIN 2024  
QY 2249 S-----RGHSFADPASNGLIEDIRKALMGSD-----DKVEDHGVMVMSOPMGVVPQTA 2297  
Db 2025 PETQAALIRGQDSTIAAS---EQVVAARILDSNDQATVNPVEFTVATPFPFSLLETSN 2080  
QY 2298 NTSVVTSGETREED---PSPH-----SGGVCKPKLISKSNSRKSPIPGQYL 2345  
Db 2081 ETDFLIGINESVEGTATYLPGRDCKMPCINGGTCTYP-----TETSYVCTCVP--GYS 2133  
QY 2346 GTERPSSVSSVHSGDYHRQTPGNWEDRPSSTGSGTQFP--YNPLTMRMLSSPTTPIACA 2404  
Db 2134 GDQCELDFTDCHS-----NPCRNGATCVDGGEN--TFRCL-----CL 2167  
QY 2405 PSAY 2408  
Db 2168 PSYV 2171

## RESULT 40

US-08-459-568-2  
; Sequence 2, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/459,568  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1706 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-459-568-2

Query Match 2.2% Score 274.5; DB 2; Length 1706;  
Best Local Similarity 18.08; Pred. No. 1.3e-08;  
Matches 365; Conservative 224; Mismatches 704; Indels 739; Gaps 81;  
QY 502 NQIARPSQEEKVEEEDKAEKTEKKEEKDEKEDSKENTKEKDIDGTABET 561

Db 144 NPEIAAAATEERASARSKRSSPKS---RRGKKSHENKNGKIRHTPTQLKASELST--- 197  
QY 562 EEREQATPRGRKTANSQGRKGRITRSMTNEAAAAAASAAAAATEPPPPPLPPPPPEISTE 621  
Db 198 -----FANRGSAGC---PKEEDERPLASA-----PEQPAPLPE----- 228  
QY 622 PVETSRWTEEMEVAKGLVHGHRNAAIAKMGVTKSEAQCKNFYFNKRRRINLNLQO 681  
Db 229 -----VG-----NQDAVPOV 238  
QY 682 HKQTSRKPREE---RDVSOCE-SVASTVSAQEDIEDASNEENPENDESEYAEVAKPSD 736  
Db 239 APLPACFPQPEVDCKQEVTDCEVNDVEEELEEELEEELEEELEGEEDGVVEADMPNES 298  
QY 737 SPENATSGNTEPAVELE---PTTETAPSTSPSLAVSTKPAEDSVETQVNDVSAETA 793  
Db 299 SAKEPEIRCEKEPEDLLEPQSMNEAREDSPTVPPHTPRAREEANGDVLETFMFC- 357  
QY 794 EOMDVDOQEHSAEESGVCDDPPATKADSDVDVVRV---PENHASKVVEGDNTERDLDRAS 850  
Db 358 -----QH-----CERKEFK-QGLERHMHHTISTINHAFKCKYCGKR----- 393  
QY 851 EKVEPRDEDLVVAQINAORPEPQSDND---SSATCSADEVDVGEPEPERQRMFMDSKPS 906  
Db 394 -----FGTOINRRRHRRHETGLKRRPSMTLOSSEDPDGGKGNVTSKDESSPP 442  
QY 907 LLNPTGSLVSSPLKPNPLDLPLQHQHRAVIPPVWVSCPTPCNIPIGTPVSGYALYORHKA 966  
Db 443 QLQDCLILNSEKTSQEVILNSFVEENG-KELHPCKYCKKVFGT---HTNMRHORR 497  
QY 967 MHE-----SALLEEORQOEQIDLECRSSTPCGTSKSPNREWEVLQAPAPHOLI 1015  
Db 498 VHERHLIPKGVYRKGGLLEEQPPAEQ---APPQNVVVPTEPEEGETDDVYIMDIS 553  
QY 1016 TNLPEGV-----RLPTTRTPRPPPLIPSSKTTVASPKSPFIMGSSISQGGTPGYLHSH 1069  
Db 554 SNISENLNYYIDGKIOTNSST-----SNCDDVEMESNSAHLGIDCLLTPTVTEITO 605  
QY 1070 N---OASYTOETPKPSVGSISLGLPROEASAKSATLYIKOEESPRSONSQPGLLVR 1125  
Db 606 NIKSTQSVTDDLLKDDSPSNCE-SKKRRTASPVLPKIKTETESDST----- 653  
QY 1126 AQHEGVVRGTAGAIQEGSITRTGTPSKISVESIPSLRGSITQGTALPQTGIPTEALVKG 1185  
Db 654 -----APSCSL-----LPL----- 663  
QY 1186 SISRMPIEDSSPEKGREAAKSHVIEYEGKSHILSYDNKINAREGTRSPRTAHEISLKR 1245  
Db 664 STSTAENVVSFHKEG-----VYLSKSLKQL-----QTDKLTLP 699  
QY 1246 SYESVEGNIKOGMSMRESFVSNAPLEGILICRALPGRSPHSDLKERVLSGSMQGTPTATT 1305  
Db 700 GFSAAE-----IPKLGVCASAP-----ASMLP-----VTS 725  
QY 1306 ESFEDGLKVPKQIKRESPPIRAFEGAITKGPYDGITTIKEMGRSITHEIPRODILTQSR 1365  
Db 726 SREKERTSSPPSQHSPALRDF-----GKPDG-----KAA 757  
QY 1366 KPEVVOSTRPIEGSISQGTPIKFDNNSGOSAIKHNKYSKITGSKLSRGMPPLEIYPE 1425  
Db 758 WTDVTLTSKCP-----KLESRSDDSPAWS-----LSGRDERETGSPPC----- 794  
QY 1426 NIKVVERKYEVDKAGE---TVRSRHTSVVSGSPSVLRSTLHEAPKAQISPGIYDTSARR 1483  
Db 795 -----FDEYKISKWEAASSTFSVCNOOP-----LDLSSG----- 825  
QY 1484 TPVSQNTMSRGSPPMNRSTSDVTIPPKNKSTNHERKSTLTPTQRESIPAKSPVPG---VDP 1540  
Db 826 -----KQKSEGTCKTPVPWESVLDL 845  
QY 1541 VVSHSPDPHHRGSTAGEVYVWHLPTQLDPAMPFHALDPAAAYLFQRLQSPTGPGSPQ 1600







Db 554 SNISENLNYYIDGKIQNSST-----SNCDVIESNSAHLYGIDCLLTPVVEITQ 605  
Qy 1070 N-----QASYTOETPKPSVGSISLGLPRQESAKSATLPYIKOEFEFSRSONSQEGLIVR 1125  
Db 606 NIKSTQVSTDDLLKDSPTSINCE-SKKRRTASPVLPKIKTETESDST-----653  
Qy 1126 AQEGVVRGTAGATQESITRGTTPTSISVESIPSLRGSTQGTTPALPOTQIPEALVKG 1185  
Db 654 -----APSCSL-----LPL-----663  
Qy 1186 SISRMPIEDSSPEKREAAKSHVIEGKSHILSYDNIKNAREGTRPTAHEISLKR 1245  
Db 664 SISTAEVVSFHKEG-----VYLSSKLQLL-----QDKLTLTPA 699  
Qy 1246 SYESVEGNIQOGMWRSPVSAPLEGITCRALPRGSPHSLDKERTVLGSGIMQGTTPRATT 1305  
Db 700 GFSAAE-----IPKLGPCASAP-----ASMLP-----VTS 725  
Qy 1306 ESFEDGLKYKQIKRESPPPRAFEGATKGPYDGIITIKEMGRSIIHEIPQDILTOESR 1365  
Db 726 SRFKRTSSPPSPQHSPPALRDF-----GKPDG-----KAA 757  
Qy 1366 KTEPVOSTRPIIEGSIQGPPIKFDNNSGOSAIKHNKSLITGPKSLRGMPPLEIYPE 1425  
Db 758 WTDVILTSKKP-----KLEKSDSPAWS-----LSGRDERETGSPC-----794  
Qy 1426 NIKVVERKGYEDVKAGE--TVRSRHTSVVSGSPVLRLTHEAPKAQLSPGIDYDTSARR 1483  
Db 795 -----FDEYKISKWEAASSTFSSVCNQOP-----LDLSSGV-----825  
Qy 1484 TPVSYQNTMSRGSPMMARTSDVTTPPNKSTNHERKSLTPTQRESIPAKSPVPG---VDP 1540  
Db 826 -----LQKVLLNEX-NGVSLTETPTTEPTRSPSPCKSPDTQDPPEL-GPDSKCVPT 928  
Qy 1601 YQLYAMENTROTINDXITSQOMQVNLRPDVARGLSPREQPLGLPYPATRIIIDLNNPP 1660  
Db 846 SVHKKPCD-----SEGKEFENHL-----AQPA-----KKKPTCM---878  
Qy 1661 TILVPHCGTSTPMWRITYIPGQITFP--RPNYASMSGPHGTHLAAAASAEERER 1718  
Db 929 AESPPVVGSPSPLOQASLSSG---QLPPLLTPTERPSPCPVLTATP-----977  
Qy 1719 EREXERERITAAASDLYLRPGSEQCRGSHGVYRSPSVRTOETMLQORPSVQGT 1778  
Db 978 -----PPPLLTPTVPLSH-----PS-----SDASPOCCSPFSNT 1006  
Qy 1779 NGTSVITPLDPTAQLRIMPLP-----AGGP-----SISQGLPASRYNTAADA--1820  
Db 1007 TAQSPLPLTSPVSPSPPIPPVEPLMSAASPGPTLSSSSSSSSPSSCSSTSPSP 1066  
Qy 1821 -LAALVDAASAPOMDVS-----KTESKHEAARLEENLRSAVSEQQOLEOKTLEV 1873  
Db 1067 PLSAVSSVSSGDLNLEASLPAVTEKQEESEGLKPKX-----EAPPAGGQSVVQETP--1119  
Qy 1874 EKRSVQCLYTSAPSPGKQPHSSVYSEACKDKGPPPKSYEEELRTRGTTITANFI 1933  
Db 1120 -SKNFIQNCVSPSLSKDLTKHLSVHAEEW-----PFKEFCVOL-FKVKTDLSEHRL 1172  
Qy 1934 -----DVII-----TRQIASDKDAREGSSQSSSSLSHRYETPSDAIEV 1975  
Db 1173 LHGVGNIFVCSVCKKEFAFLCNLQHQDLHPDEVCVTHHEFESGTLRPNQNTDPSKA---1229  
Qy 1976 ISPASSPAPPEKLOTQOYEVVYKQAE-----ND 2005  
Db 1230 -NVEHMPSLPEPLETSREEELNDSSEELYTIKIMASGIKTKDPDVRGLGNQHPSPFK 1288  
Qy 2006 PTROYEGLPHHYRQQQSPS-----PQ---QQLPSPSQAEQMGQVPTHLITLAD 2053  
Db 1289 PPQY-----HHRNPMGIGVATNFTTNIPQTFTTACTKCGKGVNDMPLELH-----1338

Qy 2054 HICQIITQDEFARNOVSSOTPOQPPTSTFQNSPSALVSTPVRTKTSNRYSPESQAQSVHHQ 2113  
Db 1339 HILACASASDKRTPKKNP-VPLKQTVQPKNGVV---LONSGKNAPRPMQPKRLSFN 1394  
Qy 2114 RPSKRVSPENL-----VDSKRSRSP-----GKSPERSHVS-----2143  
Db 1395 VELGKMSPNKLLKLSALKKKKNOLVQKAILQKNRAAKQADRLDTSEASHICPCYCDREFTY 1454  
Qy 2144 -----SEPYEPTISPPVPPVHEKQDSILLLSORGAEPAEORNDARSPGISYSLPS 2193  
Db 1455 IGSINKHAFSCPKPKPLSPSRKVSHS-----SKGGHASSSSDRNS-----1497  
Qy 2194 FFTKLENTSPMVKSKQEIFRKLNSGGGSDMAAAQPGTEIFNLPAVTTSGSVSRGH-2252  
Db 1498 -----SCHPRRTADTEI--KMOSTQAPLGKTRARSTGPAQASLP---SSSPSRQNV 1545  
Qy 2253 SPADPASNLGLEDIRKALMGSDFKVEDHGVMSQPMGVVPGTANTSVTSGETRRREG 2312  
Db 1546 KFAASVKS-----KKASSSL-----RNSPIRMAKITHVEG 1577  
Qy 2313 DPSPHSGVCKPKLISKNS-----RKSKSPIPGOGYLGTER 2349  
Db 1578 K-----APKAVAKSHSAQLSKSRGLHVRVOKSKAVIQSKTALASKR 1620

## RESULT 42

US-08-516-859A-2  
; Sequence 2, Application US/08516859A  
; Patent No. 6069231  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/516,859A  
; FILING DATE: 18-AUG-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/399,411  
; FILING DATE: 06-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292,683  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 1776  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1706 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-516-859A-2







APPLICANT: McKay, Ronald D.G.  
 APPLICANT: Lendahl, Urban  
 TITLE OF INVENTION: Nestin Expression As An Indicator of  
 TITLE OF INVENTION: Neuroepithelial Tumors  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: Two Militia Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02173  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/853,913  
 FILING DATE: 19920319  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/660,412  
 FILING DATE: 22-FEB-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/603,803  
 FILING DATE: 25-OCT-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/201,762  
 FILING DATE: 02-JUN-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/180,548  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Granahan, Patricia  
 REGISTRATION NUMBER: 32,227  
 REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-861-6240  
 TELEFAX: 617-861-9540  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1805 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-853-913-2

Query Match 2.18; Score 270.5; DB 1; Length 1805;  
 Best Local Similarity 19.7%; Pred. No. 2.6e-08;  
 Matches 308; Conservative 215; Mismatches 565; Indels 479; Gaps 73;

QY 21 PPHSVQVTFNTHQOEFAVDYDSSHLEYSQASQLLQOQQOQLRRR---PS---LLSE 74  
 DB 392 PTISEAPCPNNA-----EYRAQVPLS-----LQTAPEPLWLKATVPSSAILPE 438  
 QY 75 F-HPGS-----DRQERTSPFPHGSPVDHDSLEKRRPRLEQVSDSHFORVSAAV 126  
 DB 439 LEEPGGQGHGFPDLSLATLNPHPTLEAKDGESESR-----VSSIFDEGQI 491  
 QY 127 LPLVHPLLEGRASADAKPAFGKHEAPSPSPISGPGDQNASPKLSKEELIQSMD 186  
 DB 492 WELV-----EKEADIEVKVENS-----QKTQESGLDTEETQDSQG 528  
 QY 187 RVDREIAKV--EQOILKLLKKQOLEEAAKP-----PEKPVSPPPVE 229  
 DB 529 PLOKETLALGEEPLMSLK---IQNYETACKENCNSSTEGHLGTLGPEKEKQIPLKSLE 585  
 QY 230 QKRSIVQIIVDENRKAEEAHKIFEGLPKVELPLYNQPSDTKVYHENTKTNQVMKKL 289  
 DB 586 EKN-----VSEKTELENGVPLVELLK-----EDTKT-----EDQLMSPKG 623  
 QY 290 ILFFKRRNHARKQEOKICORYDQLEAWEKVDRIENPRKAKESKTREYY-----EKQ 345

DB 624 TL--KRFSSSLGKSEQEVVRPSKEGNLESWTAFKEESQPLGFGPAEDQMLERLVEKEDQS 681  
 QY 346 FPEIRKOREQO--ERFORVGORGAGLSATIARSHSEIIEIDGLSEQENNEKQMRQLSVI 403  
 DB 682 FPRSPDEEDQACRPLQKNOEPLGY-----EEAGQIILERLEKESQE-----SLR 728  
 QY 404 PPMFADAEQRRVKFTINMGLMEDPMKVYKDRQFMVMTDHEKEIFKDKFIQHPRNFGJIA 463  
 DB 729 SPEEDQEAGR---SLQKNOEPLG-YEEADQMLERLIEKE--SOESLSPENQRIQ 781  
 QY 464 SYLERKSVPCDVLVLYITTKKNENYKALVRNRYGRRGRNQIARPSQOEKVEEKEDKAE 523  
 DB 782 KPLERENQKS---LRYLEENQETFPLESRN-----QRPLRSELEVEEERQIVK 827  
 QY 524 KTEKKEEKKDEEKKD-----EKDSKENTKCKDKIDGTAETEETEREQATPRGRKTA 575  
 DB 828 PLEKVSQDSGLSLAEENVQPLRYLEEDDCINKSLLEDTHKSLGSLSDRN-----877  
 QY 576 NSGRRKGRITRSMTNEAAAAAATAEPPLPPPPPEP-----ISTEPVETSRW 628  
 DB 878 -----GDSIIIPQESQVSLRPEEEDQRIVNHLEKESQEFSS 917  
 QY 629 TEEMEYAKGLVEHGRNMAIA-----KMGYKSEAQCXNFYNYKRRINLONLLOQH 682  
 DB 918 SEEEQVMEKSL--EGENHESLSVEREDQMVESQLE-----KESQDSKSLSEDE 965  
 QY 683 KQKTSRKPREEDVYQCESVASTVSAQEDDIE-----ASNEE---ENPEDSEVEA 730  
 DB 966 SQETTF-GPLEKENAESLRLAG--ODQEEQKLEQETOQTLRAVNEQMAVSPKVDPEL 1022  
 QY 731 VKPSEDSPENATSRG--NTEPAVEL-EPTTETAPSTPSLAVPSTKPAED-----ESVETQ 783  
 DB 1023 PKPLGNDQEIARSLGKNOESLSVLEKGIETVKSLETEIIEPLETAEDLERKKSIDTQ 1082  
 QY 784 V---NDSISAFTAEOMD-----VDQ-----QEHSAEGSVCDPPPATKADSV 823  
 DB 1083 EPLWSTEVARETVPEPDEPPGSLGSDVENRETITLSLEKESQELSSLGKWNVETRVESQ 1142  
 QY 824 VEVVVPENHASKVEGDNTKERDLDRASEKVEPRDIEDLVAAQIINARP-----EPQS 875  
 DB 1143 QCLQVEGLQEEHQESLREVKQLPSSGNOQRWEDVVEGKAVQEAFLATGTGVEDKA 1202  
 QY 876 -----DNDSSATCSADEDVDG-----EPERORMPMDSKPSLLNPTGILVSS 918  
 DB 1203 ELHLRGOGGEEEAAGELLODIVGEANSLGSSSEPKQEV-PAEALDNLGGALEVPVQA 1261  
 QY 919 -----PLKPNLDLPLOQH--RAAVIPPMVS 942  
 DB 1262 SMPVETERDEDAQAGEQDSIEVTIGLEAAARTGLEQEVVVGLEDPRHFAREEAIPPSLG 1321  
 QY 943 CTPCNIPIGTPVSGYALYQRIHAKMHESALLEQORQEQIDLECRSSTSCGTSKSPNR 1002  
 DB 1322 EESVAKIAQGLEPGKPEKEAGAL-DSGILELPKTSSEA--LEQ-----GHESESM 1372  
 QY 1003 E-WEVLOPAPHQLITNLNPEGVRLPTTRP-----TRPPPLI-PSSKTTTVA 1047  
 DB 1373 EGWEEEAAS---LETSDHEGSDAPQPRPTEDEDEGAQAALATPGKLLFPCSPILTD 1429  
 QY 1048 KPSFIMGISQGTGPGTYLTSN---QASYTQET-----PKPSVSGSIS 1087  
 DB 1430 -----AHQLQPAQEGIQEAGWQPEAGSEALERVEPEFGLEGP 1469  
 QY 1088 LGLPROQESAKSATLPYIKOEESFPRSONSOPEGLLYRA-----QHEGVVGRTAG 1137  
 DB 1470 EGLQDWEEREES-----EADDLGETLPDSTPLGLYLRSPASPKWDLAGEOQLSPQDAG 1524  
 QY 1138 -----AIQESITRGTP-----SKISVESIPSLRGSITQGTPTALPOTGPTALV 1183  
 DB 1525 KEDWGPVAPRAAGLGGPPEEEEOCHGSDLSSEFEDL-----GTASLLPGVPKE--V 1576  
 QY 1184 KGSIRMP-----IEDSSPEKREE-----1203



Db 1577 ADHVQVPPVLPQACWQGGESDGFADDEESERGEEDADEGAEGAQWGWGASGGG 1636  
Qy 1204 -----AASKGHVYE--GKSGHILSYDNKNA-----REGTSRPTAHLSIKRS 1246  
Db 1637 CKVODIAQRGDPVQESVGVSG--LWDDGLRGAGAAANVPALEMVSQDSAEPSGSEE--SESAS 1693  
Qy 1247 YESVEGNIKQGMRESRPSAPLEGLICRALPRG--SPHSDLKERTVLSGSIMOGTPTAT 1304  
Db 1694 LEGEGQVTHDLDAPOE-VTSMVPG-VGDAFDIGQSPNLDSEQ---VNGKMENGLQAE 1748  
Qy 1305 TESFEDG 1311  
Db 1749 GQVVLDG 1755

## RESULT 44

US-09-041-886-23  
; Sequence 23, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Sharroz  
; TITLE OF INVENTION: Protopoctic Peptides, Dependence  
; TITLE OF INVENTION: Polypeptides and Methods of Use  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041.886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1185 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-041-886-23

Query Match 2.08; Score 258; DB 4; Length 1185;  
Best Local Similarity 19.68; Pred. No. 8.4e-08;  
Matches 297; Conservative 147; Mismatches 597; Indels 472; Gaps 71;  
Qy 538 KDEKSKENTKEDKIDGTAEETEEERQATPRGKRTANSQGR--RKGRIT--RSMTNEAA 594  
Db 4 RQNDKSMRSGRKEAPGPREELRSGRASPGGVSTSSDGKAEKSRQAKKARVEAS 63  
Qy 595 AASAAAAATEPPPLPPPEPTISTEPTVSRTEEMEVAKKGLVHGRNWAIAKMV 654  
Db 64 TPVKNQGRSE-----EISESESEETNAPKK----- 89  
Qy 655 GTKSEAOCKNFYFNYKRHNLDNLQHKOKTSRKPREERDVSO-----CESVASTVSAQ 709  
Db 90 -TKTEQELP-----RPQSPSLDSL--DGRSLNDGSSDPDRIDODNRKTSPTSISYSGVSE 142

Qy 710 EDEIEASNEENPEDSEVEAVKPSDESPENATSK---GNTEPAVELEPTTTETAPSTSPS 766  
Db 143 NDSSSSGLSGPARPYHPPLFPSPQPDSTPROPEASEFHPSTPTGYHAPMEPTT 202  
Qy 767 ----LAVPSTKPADESVETQVNDISISETABQMDVDOQEHSAEBSVCDPPPKATKADSV 822  
Db 203 SRMFOAPPAGAPP--P-----PQLYPGGTGGVLGGPMPKGGGA 240  
Qy 823 DVEVVRVENHASKVEGDNTERDLDRASEKVEPRDEDLVVAQOINAQORPEPQSDNDSSAT 882  
Db 241 -----ASSVGGPNG-----GKOHPPPTTPISSVSS 265  
Qy 883 CSADEDVDGEPERQRMFPMDSKPSLLNPTGSLVSSP-----LKNPLDLPLQ--HR 933  
Db 266 -----GASGAP-----PTKPTTTPVGGGNLPSAPPANFPHVTNLPPLPALPLNN 312  
Qy 934 AAVTPMVSTPCNIPITGTPVSGYALYQRIKAMHESALLEEQRQREQIDLECRSSTSP 993  
Db 313 ASAPPGLGAQP--LPGLHP--SPYAMGQ-----GMGGLP 343  
Qy 994 CGTSKSPNREWEVLOPAPHQLITNLPEGVRLPTTRTPRPDPL-IPSSKTTVASEKPSFI 1052  
Db 344 PGPEKGP-----TLAPSPHS-----LPPASSAPAPPMRFPYSSSSSSAAAASSS 388  
Qy 1053 MGGISQGTGTYLTSHNQASYQTETPKPSVSGISLGLPRQOESAKSATLPYIKQEEFSP 1112  
Db 389 SSSSSSSNASP--FPASQALPSYHSFPPT-----SLSVNQPKYTOPSLP--SQAVWSQ 440  
Qy 1113 RSQNSQPEGLLVRAQ--HEGVVRGTAGAIQEGSITRGTPTS-----KISVESIPSLRGS 1164  
Db 441 GPPPPPPYGRLLANSNAHPGPPPTGA---QSTAHPPVSTHHHHQOQOQOQOQOQ 496  
Qy 1165 ITQGTALPQTGIPTEALVKGISRMPIEDSSPEKREAAKSGHVIYEGKSGH--LSYD 1223  
Db 497 QHNGSGPPPPGAPPHLEGGSSHAPHYAMSPSLSLRP-----YPPGPAHLPPPHS 549  
Qy 1224 NIKNAREGTRSPRTAHEISLKRYSVEGNIKQGMRESVPS--APLEGLICALPRGS 1281  
Db 550 QVSYSOAGPNGPPV-----SSSSSSSSTSGSYPCSHSPSQG-----PQGA 592  
Qy 1282 PHSDLKERTVLSGSIMOGTPTATTESFEDGLKY-----PKQIKRESPPIRAFEGAIKRG- 1335  
Db 593 PYPPPPVPTVTTSATLSTVIATVASSPAGYKTASPPGPPYGRAPSPGAYKATATPGY 652  
Qy 1336 KPYDGIITIKEMGRSIEHPRQDILITQESRKTPEVWGOSTRII--EGSISQCTPIKFDNNS 1394  
Db 653 KP--GSPPSFTG-----TPGYRGTSPPAGPGTFKPGSP----- 685  
Qy 1395 GQSAIKHNKSLITGPSKL-----SRGMPPLEIVPENIKWVERGKYEDVKAGETVRSRHTS 1450  
Db 686 -----TVGPGPLPPAGPSGLPSLPPP-----A 708  
Qy 1451 VVSSGSPVLRLSTL-----HEAPKAQLSPG-----IYDDTSARPTVSYONTMSRG 1495  
Db 709 APASGPPLSATQIKQEPAAEYETPESPVPARSPPPKVVVDVPSHASQSAFKNKHLDRG 768  
Qy 1496 SPMMNRTSDVTIPP-----NKSTNHERK 1518  
Db 769 FNSCAR--SDLYFVPLEGSKLAKRADLVEKVRRAEORAREEKEKEREKEREKEREKE 827  
Qy 1519 STLTPTQRESIPAKSPV--PGVDVPSVSHSPFDPHHRGSTAGEVYWSHLPTOLDPAMPFHR 1576  
Db 828 RELERSVKLAQEGRAPVECPSLGVPVHRPPEP--GSAVATV-----PPYLGPDTPALR 879  
Qy 1577 ALDPAAYLFFQRLSP-----TPG-----TPSQYL--YAMENTROTILN 1615  
Db 880 TLSEYARPHV---MSPGNRNHPYVPLGAVDPLGGLGYNVPALYSSDPAAREREA--- 932  
Qy 1616 DYITTSQOMQVNLRPDVARGLSPRE--OPL--GLYPATRGIIIDLTNMPTILVPHPGGTSTP 1673  
Db 933 ---RERDLORLKGFE--VKPSELEPLHGVGP-----GLDP---PFRHGGGLALQ 975  
Qy 1674 PMDRITYPGTQITTPPRPYNSASMSFGHPHAAAAAARE-----REREREKERER 1728







```

QY 1736 LYLRPGSEQP-----GRPGSHGYVRSPSVRTQETML-----QORPSVFOG----- 1777
Db 1453 QSLNPASFSLATQAKP-----EKSTLQRPQETVIRELQPOQOQPTIERRDLQYITIS 1506
QY 1778 -----TNGTSVITP-----LDPTAQLRIMPLPAGGPSISQGL-----PASRYNTAA 1818
Db 1507 KEELSSGDS-LSPDPWKRDRAREKLEKQOQMHIVDM-----LSKEIHELQNKGDRTAES 1559
QY 1819 DALAALYDAASAFQMDVSKTESKHE-----AARLEENLRSAAYSEQOQL 1866
Db 1560 DRLRLKMLEWQFQRLQESKQKDEDEDDDDVDTMLIMORLEAERRAR--LQDEERRR 1617
QY 1867 EOKTLEVEKRSVQCLYTSSAPPSGKPPHSSVYSEAGKDKGPPPKSRYEELTRGKTT 1926
Db 1618 QQLEEMKREVE-----DRVQDEEG-----RHQEEERVK----- 1648
QY 1927 ITAANFIDVITITQIASDKDARERGSSSSSLSSHRYETPSDA-----IEVISPA-SS 1981
Db 1649 -----RDAEKKRQEEGYYSRLEAERRRQHEEAARRLLEPEEPGLSR 1690
QY 1982 PAPPOEKLQTYQPEVVKANQAEENDPTROYEGPLHHYRPOQESPSPO-----QQLPSS 2033
Db 1691 PPLFPD-----YEPQSOSAPSPAPPPPPQORNASYLKTQVLSPD 1728
QY 2034 SOAEGMGQVPRTHRLITLADHICQIITQDFARNQVSSQTPPOOQPTSTFQNSPSALVSTPV 2093
Db 1729 SLF-----TAKFVAYDD-----DDEENYV-----PAGP-----NSYSGSAGTTA 1763
QY 2094 RTKTSNRYSPESQAQSVHHQRPQSGKRVSPENLVDKSR 2129
Db 1764 GTYDAPRDTREKLSRSQDADLPGSSGAPENLTFRER 1799

```

Search completed: September 8, 2001, 14:42:40  
Job time: 447 sec